A MOLECULAR CHARACTERIZATION OF BREAST CANCER PROGRESSION, INCLUDING ONSET, TUMOR GROWTH AND METASTASIS

A Dissertation Presented to
the Watson School of Biological Sciences
at Cold Spring Harbor Laboratory

in Partial Fulfillment of the Requirements for
The Degree of Doctor of Philosophy

by

Elvin Wagenblast

Cold Spring Harbor Laboratory
July 2014

Table of Contents

List of Tables

Chapter 2: Transcriptional framework of ductal carcinoma in situ (DCIS)

Chapter 3: Functional genomics of disseminated and circulating tumor cells in a breast cancer mouse model

Chapter 4: A model of tumor heterogeneity reveals vascular mimicry as a driver of metastasis

List of Figures

Chapter 2: Transcriptional framework of ductal carcinoma in situ (DCIS)

Chapter 3: Functional genomics of disseminated and circulating tumor cells in a breast cancer mouse model

Chapter 4: A model of tumor heterogeneity reveals vascular mimicry as a driver of metastasis

Abbreviations:

AML: acute myeloid leukemia
CGH: comparative genomic hybridization
CTCs: circulating tumor cells
DCIS: ductal carcinoma in situ
DTCs: disseminated tumor cells
ECM: extracellular matrix
EMT: epithelial mesenchymal transition
ER: estrogen receptor
ESR1: estrogen receptor 1
FACS: fluorescence activated cell sorting
FDA: Food and Drug Administration
FISH: fluorescence in situ hybridization
H&E: hematoxylin and eosin
IDC: Invasive ductal carcinoma
LN: lymph node
MHC: major histocompatibility complex
NGS: next generation sequencing
PR: progesterone receptor
RBC: red blood cell
RPM: reads per million per nucleotide
ROMA: representational oligonucleotide microarray analysis
SEREX: serological analysis of recombinant tumor cDNA expression libraries
SNPs: single-nucleotide polymorphism
SNS: single-nucleus sequencing
SNV: single nucleotide variation
SNVs: single-nucleotide variants
VM: Vascular mimicry
WGA: whole genome amplification

Acknowledgments

I would like thank Greg Hannon for his great support for the past 6 years. I am very grateful for the opportunity to work with him and to benefit from his vast scientific knowledge. I am very fortunate that he gave me the freedom to tackle different projects and provided me the resources and manpower to accomplish these tasks.

I would also like to thank Simon Knott. He has been instrumental for the heterogeneity project and it has been a great experience to work with him. His scientific expertise and motivation was invaluable for the project.

I would like to thank the rest of my committee, Jan Witkowski, Marja Timmermans, Scott Lowe and Mikala Egeblad. You always had words of encouragement and your guidance kept moving the projects along. Thank you to Charles Sawyers for accepting to take the role as my external examiner.

In particular, I would like to acknowledge the contributions of Mar Soto, Sara Gutiérrez-Ángel, Christina Hartl, Alissa Williams, Annika Gable and Rebecca Berrens for their vital help at the heterogeneity, DCIS and/or CTC project. Without their dedicated work, all of these projects would not be at the level they are now. It has been an amazing opportunity to mentor these students and this experience has really shaped me, not only as a scientist. In addition, I would like to thank those who contributed work and suggestions including Ashley Maceli, Nicolas Erard, Emily Hodges, Emily Lee, Ken Chang and Maria Mosquera. I wish to acknowledge our collaborators Chuck Harrell and Charles Perou from the University of North Carolina and John Wilkinson from the University of Michigan. I also would like to thank the CSHL Shared Resources, especially the Animal Facility with Michael Cahn, Jodi Coblentz and Lisa Bianco, the Flow Cytometer Facility with Jordan Ratcliff and Pam Moody and the Histology Facility with Denise Hoppe, Aigoul Nourjanova and Raisa Puzis.

Thank you to Jo Leonardo and Sabrina Boettcher for making sure that the lab runs smoothly. I would also like to thank the administrative staff at the Watson School including Alex Gann, Leemor Joshua-Tor, Alyson Kass-Eisler, Kim Geer, Dawn Meehan, Uwe Hilgert, Keisha John, Kim Creteur and Carrie Cowan. I would like to acknowledge the Watson School for providing an excellent PhD graduate program.

I would like to thank the Starr Centennial and the Boehringer Ingelheim Fonds for their generosity in providing funds for my PhD.

I would like to thank my parents and my family for their continuous support and encouragement. Thank you to Jin Gu. I am immensely grateful for all your love.

I would like to thank all the members of the Hannon lab, past, present and future. I have greatly enjoyed my time with all of you. I have been blessed with some wonderful friends at CSHL. Thank you to Ben Czech, Tim Kees, Stefan Klinger, Timour Baslan, Max Doerfel and Simon Knott for making it fun at work. Lastly, I would to thank everyone at Cold Spring Harbor Laboratory for providing an amazing environment to work in.

Of making many books there is no end
Ecclesiastes 12:12

Summary

Breast cancer is the most common cancer among women in the United States, except skin cancer. 12% of women will develop invasive breast cancer during their lifetime. We sought to understand the transcriptional framework of breast cancer disease progression, including on-set, tumor growth and metastasis. Ductal carcinoma in situ (DCIS) is referred to as stage 0 breast cancer and represents an initial proliferation of cancer cells within the breast ducts. In order to understand how DCIS progresses to invasive ductal carcinoma (IDC), we combined laser capture microscopy on human patient samples and RNAseq technology to profile their respective gene expression profiles. We completed hierarchical clustering of these samples, classified DCIS and IDC subtypes and identified differentially expressed genes that might serve as biomarkers. Next, we used mouse models of breast cancer metastasis, to assess the gene expression profiles of circulating tumor cells (CTCs) and metastatic tumor cells in order to identify genes that are regulators of the metastatic cascade. We identified Ptpn22, a protein tyrosine phosphatase, as a metastatic inducer. Knock-down of Ptpn22 decreased the lung metastatic burden in mice, due to an overall decrease in the number of CTCs in the blood circulation. Lastly, we developed a mouse model to assess clonal heterogeneity of primary tumor growth and metastasis. We identified clones that showed distinct specialization such as dominating in the primary tumor or metastasizing to the lung. Using this heterogeneity model, we identified Serpine2 and Slpi, two secreted protease inhibitors that are necessary and sufficient to program cancer cells for vascular mimicry. We think that vascular mimicry is one of the major drivers of metastatic progression. In summary, we identified multiple genes that might be relevant to human breast cancer progression and metastasis.

Chapter 1: Introduction

Breast cancer

Breast cancer initiates in the breast tissue, which is composed of glands for milk production, called lobules, and ducts, which connect the lobules to the nipple. In addition, there is connective, fat and lymphatic tissue. Cancers that occur within the ducts are known as ductal carcinomas and those that originate from the lobules are known as lobular carcinomas. According to the American Cancer Society, breast cancer is the most common cancer in females in western countries with one in eight women developing it (DeSantis et al. 2014). Approximately 60,000 new diagnoses of non-invasive *in situ* breast cancer and 200,000 new cases of invasive breast cancer are expected among women in the United States in 2013 (Siegel et al. 2013). In addition, approximately 40,000 breast cancer deaths are expected among US women in the United States (Siegel et al. 2013). Nowadays, breast cancer accounts for 29% of all new cases of cancers and is the number two leading cause of cancer deaths (15% of all estimated cancer deaths), just behind lung and bronchus cancer in the United States (Siegel et al. 2014).

Ductal Carcinoma In Situ (DCIS)

Ductal carcinoma in situ (DCIS) is a spectrum of abnormalities that represent regions of proliferating neoplastic epithelial cells that line the breast ducts and have not yet invaded the ductal basement membrane. DCIS is the most common type of *in situ* breast cancer, representing around 80% of the *in situ* cases in the United States (Barnes et al. 2012; Virnig et al. 2010). It is believed that DCIS is the precursor to invasive ductal carcinoma (IDC). It is estimated that at least one-third of the cases diagnosed with DCIS will advance to invasive cancer if left untreated (Allred 2010).

The pathways of progression from DCIS to IDC are not well understood. The standard treatment of DCIS is breast-conserving surgery followed by radiation therapy to the intact breast. Approximately 20% of women who where diagnosed with DCIS and received treatment had local recurrence within 15 years, half of these cases were invasive carcinoma (Donker et al. 2013). On the other hand, those patients who were diagnosed with DCIS and were left untreated, approximately 35% of the patients developed IDC (Meyerson et al. 2011). Interestingly, the local recurrence is significantly higher in women under 40 years at approximately 30%, whereas women over 40 years have a local recurrence of approximately 10% (Vidali et al. 2012). Thus, it is important to dissect the molecular pathways of DCIS in order to understand why some women progress to invasive carcinoma, relapse with DICS or IDC, or why young age is a big risk factor for local recurrence. By identifying molecular determinants of DCIS and IDC and correlating them with clinical annotations, we will be able to stratify patient outcome, and ultimately avoid over-treatment of patients diagnosed with DCIS.

The DCIS incidence rate dramatically increased in the United States in the last 40 years, from 1.87 per 100,000 women (1973 – 75) to 32.5 per 100,000 women in 2004, due to the widespread use of mammography. However, the number of new cases of invasive carcinomas has not decreased in the last decades (Virnig et al. 2010). One explanation might be that the outcome of DCIS is pre-programmed or that the treatment itself might promote progression to invasive carcinoma due to the use of ionizing radiation during therapy (Broeks et al. 2007). Over half of the patients diagnosed with DCIS will never develop invasive carcinoma, but all of these patients will most likely opt for treatment to avoid the risk. With very little knowledge about the causes of the transition between DCIS and IDC, there is an overwhelming need to identify, which DCIS lesions will develop into invasive

carcinoma. In order to identify these molecular determinants, we generated in-depth and unbiased gene profiling of multiple single DCIS and IDC lesions using RNAseq to probe their respective gene expression.

Breast cancer metastasis

The metastatic form of breast cancer, which can spread to lung, liver, bone, lymph nodes, and brain accounts for 90% of breast cancer deaths (Cifuentes and Pickren 1979; Steeg and Theodorescu 2008; Talmadge and Fidler 2010). Treatment of breast cancer often involves surgical removal of cancerous tissue, but, even after successful operation, recurrence rates are high (Karrison et al. 1999). While different approaches such as surgery, hormone therapy, chemotherapy and/or radiation therapy are used to eliminate metastatic cells, dormant solitary cancer cells can cause outbreaks after latency periods of 10 to 30 years (Aguirre-Ghiso 2007; Weckermann et al. 2001; Pfitzenmaier et al. 2006). To date, much research has focused on understanding the origins and mechanisms of breast cancer. However, the biology of disseminated tumor cells and their metastases are still poorly understood.

In order to understand the mechanism of tumor metastasis, it is important to investigate every stage of the metastatic cascade. Metastasis is a stepwise mechanism starting with tumor cells invading through the basement membrane, following intravasation into the blood or lymphatic system and its subsequent survival, then extravasation into distant organ sites, and finally colonization into macro-metastases (Coomber et al. 2003). In order to metastasize, it is essential for cancer cells to complete all the stages described above. Even though the different steps of the cascade are nicely defined, the cellular and molecular mechanisms of dissemination are still poorly understood (Fein and Egeblad 2013). Advances

in next-generation sequencing (NGS) and low-input RNA library generation have created new possibilities to understand this metastatic cascade. RNAseq libraries can now be generated from very small amounts of cells, down to single-cell levels. Because of this, it is possible to dissect the dissemination process by performing transcriptome profiling of primary tumor cells, circulating tumor cells (CTCs) in the blood, and metastatic tumor cells at distant organ sites. Because the invasive behavior coincides with altered gene expression, by comparing the expression signature between these different compartments, we can identify genes that are responsible for certain steps of the metastatic cascade.

Although tumors originate from a clonal origin, a large degree of heterogeneity can be found in tumors. These cells vary in their morphological as well as their molecular features (Shipitsin et al. 2007; Fisher et al. 2013). Clones that have been isolated from a murine melanoma cell line and injected intravenously into syngeneic mice, displayed different metastatic potentials to the lung (Fidler and Kripke 1977). This indicates that the primary tumor is heterogeneous and more importantly that it contains different cell variants of varying metastatic capabilities. For this reason, it will be important to identify and study these individual cell variants as opposed to the bulk tumor population, since these cells will allow a greater insight into the biology of the metastatic process.

Circulating tumor cells (CTCs)

Metastatic spread is a complex process initiated by the dissemination of the primary tumor, subsequent seeding and engraftment of malignant cells at distant organ sites. In many cancer types, metastasis occurs through the lymphatic system, however in breast cancer, it has been found that 20 to 30% of patients have no lymph node metastases. This suggests that primary tumor cells can bypass the lymphatic system and can directly

disseminate through the blood stream, as circulating tumor cells (CTCs) (Pantel and Brakenhoff 2004). Another explanation is that the lymph node microenvironment might not be favorable for some tumor cells. CTCs are thought to be arrested in the G_0 phase as demonstrated by an absence of the proliferation marker Ki67 (Pantel et al. 1993). Because of this, CTCs might be resistant against cytotoxic therapeutic agents (Braun et al. 2000). The presence of CTCs is also associated with a significant lower survival rate in breast cancer patients. The median overall survival was 18 months for CTC-positive patients, compared to 27 months for CTC-negative patients using the Food and Drug Administration (FDA)-approved CellSearch system for CTC enumeration (Zhang et al. 2012). Thus, much research is currently focused into developing tools to isolate and analyze CTC for prognostic and therapeutic purposes. It is believed that cancer outcome can be analyzed more precisely by molecular profiling of CTCs. However, whole-genome expression analysis of CTCs has not yet seen widespread use (Krebs et al. 2014). To date, only three studies were carried out with whole-genome transcriptome profiling, namely in prostate, melanoma and breast CTCs (Yu et al. 2013; Cann et al. 2012; Ramsköld et al. 2012).

Sensitive methods have been developed to detect and isolate CTCs in the peripheral blood. CTCs can be distinguished and enriched from surrounding blood cells either by physical properties such as size and density or biological properties such as the expression of epithelial markers like EpCAM or cytokeratins (Alix-Panabières and Pantel 2014). The disadvantage of using epithelial markers to enrich for CTCs is that this might miss cancer cells that have undergone epithelial mesenchymal transition (EMT). Indeed, many CTC populations are enriched for mesenchymal cells in breast cancer (Yu et al. 2013). Thus, it will be important to include these cells in the isolation process.

Although systemic anti-cancer therapies such as chemotherapy are designed to eliminate metastatic cells, the current classification for treatment options is most often performed on the primary tumor. Increasing evidence shows that the phenotype and genotype of primary and metastatic tumor cells are discordant. For example, Krishnamurthy et al. reported a discordance in Her2 status between CTCs and primary tumors of early stage breast cancer by fluorescence in situ hybridization (FISH) (Krishnamurthy et al. 2013). 21% of patients showed acquired Her2 gene amplification in CTCs that had Her2 negative early stage breast cancer. Thus, it is possible that a very small fraction of primary tumor cells are Her2 positive and that these cells remained undetected by FISH. These cells might represent a potent metastatic subpopulation.

Technological advances have enabled detection and characterization of CTCs from various solid tumors. So far, individual markers such as Her2 are used to analyze CTC populations. Currently, much effort is concentrated to carry out whole genome transcriptome analysis of CTCs using next generation sequencing. These studies will shed light into the genetic heterogeneity during metastasis and will offer new insights into the metastatic progression.

Epithelial Mesenchymal Transition (EMT)

It has been proposed by some investigators that in order for tumor cells to infiltrate the blood stream, they undergo an epithelial to mesenchymal transition (EMT). For this, epithelial cells can activate a transitory EMT program and convert into a mesenchymal state with increased migratory and invasive capabilities. It is thought that this stage is fluid, with many cells expressing epithelial and mesenchymal markers at the same time (Yu et al. 2013).

Fred Lillie first described the difference between epithelial and mesenchymal cells in 1908. However, the mechanism of EMT was not defined until 1982 (Greenburg and Hay 1982). EMT plays an important role in embryonic development during the reorganization of germ layers (Shook and Keller 2003), but has only been recognized to play an important role in tumor progression and metastasis in the last decade (Thiery 2002; Thiery et al. 2009).

Many cell lines and mouse tumor models have demonstrated the importance of EMT during the metastatic cascade (Giampieri et al. 2009; Yang et al. 2004; Mani et al. 2008). Nevertheless, the EMT process in human cancers is difficult to identify since the cells that undergo EMT share many morphological and molecular features of mesenchymal stromal cells such as fibroblasts (Thiery 2002). Several EMT genes such as Twist, Snail and Slug have been described to have a role in metastasis (Emadi Baygi et al. 2010; Hotz et al. 2007). Therefore, by targeting EMT genes it may be possible to inhibit the EMT program and thereby preventing the metastatic progression.

Tumor Microenvironment

Solid tumors do not only represent a random assembly of cells but can be thought of as a separate organ. All of the cell types comprising an organ can also be found in tumors. The different tumor niches can be defined by (1) members of the extracellular matrix (ECM) (2) stromal cells including mesenchymal supporting cells, cardiovascular cells as well as cells of the immune system and (3) peptide factors covering growth factors, chemokines, cytokines and metabolites (Egeblad et al. 2010). All of these different factors play an important role in the tumor microenvironment, which in turn have a great impact on the regulation of cellular movement and assembly during tumorigenesis (Bissell and Radisky 2001; Wiseman and Werb 2002; Bissell and Labarge 2005). While some of these factors have

tumor-promoting effects, others have strong inhibitory effects on tumor growth. Due to their rigid barrier, normal epithelium and fibroblasts have a tumor inhibiting effect in hyperplastic tissues. Nevertheless, in the process of tumor development carcinoma-associated fibroblasts support tumor growth and have large effects on changes in the ECM (Orimo et al. 2005).

Cells of the immune system play a role in tumor-inhibition as well as in tumor-promotion. Inflammatory cells like neutrophils and macrophages are recruited to the primary tumor due to their innate and adaptive immune response (Fridlender et al. 2009). Based on their polarization they are either pro- or anti-tumorigenic. Mast cells have been described to promote angiogenesis and thereby promoting tumor development (Coussens et al. 1999). Most dendritic cells support immunosurveillance and are therefore tumor-inhibiting factors (Knight et al. 1985; Mayordomo et al. 1995). However, natural killer cells as well T-cells can either inhibit or promote tumor growth (DeNardo et al. 2009; Casares et al. 2003; Curiel et al. 2004; Girardi et al. 2001). Since immunoglobulins, produced by B cells, initiate the inflammatory response they rather inhibit tumor development (Inoue et al. 2006). The metastatic process is also strongly connected to the tumor microenvironment. Fibronectin is strongly expressed on vascular leakage sites promoting the exit of tumor cells from the vasculature (Kaplan et al. 2005). This further promotes the formation of metastases. It is thought that activated platelets cover CTCs to protect them during their travel through the blood vessels as well as their adherence to endothelial walls in distant organ sites (Camerer et al. 2004; Nieswandt et al. 1999).

It is still unknown what role the microenvironment plays in the transition from DCIS to IDC. There is evidence that stromal cells might play a role in this transition, since a macrophage response signature is associated with a higher grade of DCIS, decreased

expression of estrogen receptor (ER) and progesterone receptor (PR) and an increased rate of p53 mutations (Sharma et al. 2010). In addition, periductal macrophages and myofibroblasts have been associated with the initial steps of invasive carcinomas by up-regulating the expression of the plasminogen activation system and MMP-13 (Nielsen et al. 2007; Fakhrejahani and Toi 2012). These studies indicate that the stromal compartment might possess a role in the transition from DCIS to IDC and thus might be useful as predictive markers.

4T1 murine breast cancer mouse model

The 4T1 breast cancer cell line series (4T1, 66cl4, 4T07, 168FARN and 67NR) have been isolated from a single primary tumor of a BALB/c mouse. These mice have been infected with the mouse mammary tumor virus (MMTV), which has properties similar to other RNA tumor viruses and is present in mothers' milk. By expression of key host genes in the mammary stem cell pre-neoplastic epithelium, a tumorogenic phenotype is induced. Interestingly, the 4T1 cell lines differ in their metastatic potential (Aslakson and Miller 1992). Additionally a difference in the *in vitro* growth rate, sensitivity to chemotherapeutics as well as MMTV expression distinguishes the different cell lines (Pulaski and Ostrand-Rosenberg 1998; Ostrand-Rosenberg et al. 1998). The 4T1 cell line resembles most reliably the human stage IV breast cancer in respect to growth kinetics, pathology, invasiveness, poor immunogenic phenotype as well as spontaneous metastasis formation (Pulaski et al. 2000).

While 4T1 and 66cl4 cell lines lead to lung metastasis upon orthotopical transplantation of cells into the mammary gland of BALB/c mice, no overt metastases can be found in mice injected with 4T07, 168FARN and 67NR cell lines. The 4T1 cell line is highly aggressive and metastasizes spontaneously to lung, liver, bone and brain in BALB/c

animals. The 66cl4 cells mostly metastasize through the lymphatic system to the lung and these cells were mutagenized with EMS. The 4T07 and 168FARN cells were also mutagenized with EMS and the 67NR cell line was not treated with any mutagen (Miller et al. 1983b; Dexter et al. 1978; Aslakson and Miller 1992; Miller et al. 1986).

Tumor heterogeneity

In the early 1800s Rudolf Virchow was one of the first people to observe morphological heterogeneity in tumor tissues (Brown and Fee 2006). The first concrete evidence for tumor heterogeneity was discovered in the 1950s, when they identified different cell populations throughout the tumor (MAKINO 1956). Studies in the 1970 showed evidence for distinct subpopulations of tumor cells, which differed in tumorigenicity, resistance to therapy and metastatic potential (Fidler and Kripke 1977). This was perceived with skepticism, since it was initially thought that tumors originated from a single mutated cell and thus should be monoclonal (Fialkow 1979). In 1976, Peter Nowell proposed a clonal evolutionary model of cancer development and subsequently the concept of tumor heterogeneity received broader acceptance (Nowell 1976). Nowadays we know that tumor heterogeneity is a consequence of somatic evolution of cancer cells caused by stepwise clonal expansions. Single tumor cells undergo developmental changes during their growth by acquiring random mutations. These tumor cells undergo positive selection for advantageous mutations and negative selection for harmful mutations. In the case of an evolutionary advantage, the generated cells can outcompete their less fit parental and sister clones (Hanahan and Weinberg 2000). Because of this, genetically related clones with slightly different genotypes and phenotypes can cohabit in the same tumor (Shipitsin et al. 2007). This polyclonal composition is advantageous for the tumor, since the molecular

heterogeneity increases the pool of genetic variants that undergo selection and thus increases the chance of survival for one or few subclones (Fisher et al. 2013). With different subpopulations, together they can cope with a wider variety of environmental factors. Overall, this can results in the generation of different superior subpopulations in one tumor.

The ability of tumor cells to accumulate DNA mutations is greatly increased compared to normal cells. This is known as genomic instability and is one of the hallmarks of cancer. Other hallmarks include sustained proliferative signals, ability to evade growth suppressors, replicative immortality, resistance to cell death, immune evasion, promotion of inflammation, angiogenic potential, increased motility, metastatic ability, and altered cellular metabolism (Hanahan and Weinberg 2011). Heterogeneity can be separated in a mosaic pattern throughout the tumor or in different subpopulations that are regionally separated. With different environments generated by hypoxia, angiogenesis, necrosis and growth, it would be very surprising if all tumor cells were identical (Gerlinger and Swanton 2010). Thus, clonal composition can not only differ in the primary tumor, but between the primary tumor and metastases or tumors of the same histopathological subtype (Komaki et al. 2006).

The development of cellular heterogeneity within a tumor is a complex process. Driver and passenger mutations are acquired due to the increased instability of the genome. On one hand, driver mutations increase the fitness of a clonal cell and confers a selective advantage, passenger mutations have no negative effects on the proliferation potential. Currently, only driver mutations stand in the central focus of research, when trying to understand the genetic basis of cancer.

Intratumor heterogeneity poses a great challenge for personalized medicine and selection of treatment. Thus it is important to develop methods to detect tumor

heterogeneity in order to identify and model therapeutic responses and disease outcome. New technologies are being used and developed in order to identify tumor heterogeneity.

Approaches to identify tumor heterogeneity

In order to gain a deep understanding into the architecture of cancers, genomic analyses are necessary. Early studies based on Sanger sequencing were insufficient to analyze tumor heterogeneity because sequencing depth was limited and rare subpopulations could not be detected. In addition, extraction of tumor samples always yield stromal cells, such as fibroblasts or immune cells. This can complicate genomic DNA isolation of tumor cells and can provide false positives in the sequencing results. In addition, analyzing bulk DNA from tumors can lead to loss of detecting cellular heterogeneity at a single cell level. Nevertheless, through the development of novel technologies within the last few years, we are able to examine the clonal composition of cancers and tumor heterogeneity at a high resolution.

Allele frequency quantification

The first approach to study clonal evolution in cancer using next-generation sequencing (NGS) was developed by Elaine Mardis in 2009 (Mardis et al. 2009). The technology is based on the identification of somatic mutations and subsequent calculations of allele frequency. First, the genomes of an acute myeloid leukemia (AML) tumor and normal skin were sequenced using NGS. Then, single-nucleotide variants (SNVs) were detected and compared to single-nucleotide polymorphisms (SNPs). After validation, deep sequencing reads were obtained of each somatic mutation in order to calculate the allele

frequency of variants in the tumor cell population. With this, groups of mutant allele frequency could be identified and this was compared to normal skin samples.

Using the same approach, the same group also analyzed the heterogeneity of tumor progression in basal breast cancer by looking at the allele frequency of the primary tumor and metastases (Ding et al. 2010). One of the questions they wanted to address was whether the metastasis process requires specialized cells generated by the primary tumor or whether the necessary mutations occur only after the cell has arrived at the metastatic site. For this, genomic DNA was extracted from peripheral blood, the primary tumor, brain metastasis and an early passage xenograft. Most of the identified mutations were found throughout all cancer samples, but some mutations were highly enriched in the metastases and the xenograft. Interestingly, these genes were involved in migration, invasion and tumorigenesis (Chan et al. 2008). Several mutations were also only found in the metastases, but not detected in the primary or xenograft tumor genomes. The authors were able to conclude a hierarchical model of genomic changes with mutations occurring at different levels during disease progression, constituting to a considerable genetic heterogeneity of the primary tumor. Mutation frequencies were lower in the metastases and xenograft compared to the primary tumor, indicating that only certain cells with distinct mutations were able to form metastases or grow as a xenograft. Interestingly, mutation frequencies often overlapped between these metastases and xenografts, highlighting a similar selection process between both environments. Thus, in summary, the potential of forming metastases seems to already be embedded in the cancer cells when they separate from the primary tumor. However, clonal evolution continues at the site of metastases and contributes to the overall metastatic phenotype. Using this approach, the group also analyzed the clonal composition in relapsed AML (Ding et al. 2012).

Overall, these studies give strong evidence to the concept of tumor heterogeneity, which is a step-wise accumulation of mutations. Limitations of this method are that it is not able to detect rare mutations at a single-cell level, due to low sequencing coverage of the whole genome. Furthermore, this method is unable to uncover more complex clonal relations. Lastly, it is impossible to do functional studies on these subpopulations, because they can not be easily isolated.

Single nucleus sequencing

Another approach to trace tumor development by studying mutations in different subpopulations of tumor cells was established by Nicholas Navin and James Hicks (Navin et al. 2010). Their goal was to construct a phylogenetic lineage of cancer progression by comparing data from divergent clones of a primary breast tumor and its metastases. For this, they macro-dissected solid tumors into sectors and then isolated nuclei from each sample. Fluorescence activated cell sorting (FACS) was used to sort nuclei according to differences in their genomic DNA content as measured by DAPI intensity (Corver et al. 2008). DNA was isolated from the flow-sorted nuclei and then used for whole-genome copy number variation analysis by representational oligonucleotide microarray analysis (ROMA) (Lucito et al. 2003). This form of comparative genomic hybridization (CGH) allows for analysis of copy number profiles and chromosome breakpoint patterns. Thus, samples of different tumor sectors and tumor subpopulations based on DNA content were compared to assess heterogeneity throughout the tumor.

Using this method, Navin and Hicks were able to see that slightly more than half of the analyzed breast tumors were classified as polygenomic and displayed considerable heterogeneity. The separation of subpopulations based on DNA content matched the

profiles that were seen using the CGH analysis. Within one tumor most chromosome variations were very similar between the subpopulations, which strongly indicate that all subpopulations had a common clonal origin. On the other hand, some tumors contained subpopulations that shared almost no chromosomal breakpoints. Overall, they were able to conclude that mutations are accumulated as the tumor progresses and genomic heterogeneity can serve as a permanent record of mutations, unless a whole chromosome is lost. A major drawback is the limitation of analyzing mixtures of cell populations. Different clones can only be separated based on regional sectors, surface markers or ploidy to conduct genome-wide comparisons.

In order to overcome this limitation, Navin and Hicks used single-nucleus sequencing (SNS) in their subsequent study (Navin et al. 2011). For this, nuclei were isolated from whole tumor samples by FACS. Genomic DNA from single cells was amplified by means of whole genome amplification (WGA), sequenced using NGS and copy number variations analyzed. This allowed for a closer examination of individual cells of different tumor subpopulations and indeed revealed that each subpopulation is comprised of cells with highly similar copy number profiles. Although every subpopulation shared attributes with others, they still showed distinct genetic profiles. This supports the notion that all subpopulations diverged from a common ancestor at an early stage of tumor development. They were also able to analyze the heterogenic properties of matched primary tumor and metastases. Interestingly, they saw that the metastatic tumor was seeded by only one cell of the primary tumor, which developed with little further evolution. Although similar, the metastatic cells clearly separated from the primary tumor cells, indicating that there was no intermixing between the two populations after formation of the metastasis. Analogous

results have been reported in other studies (Liu et al. 2009; Hovey et al. 1998; Israeli et al. 2004; Ding et al. 2010).

The study by Navin and Hicks strongly support the clonal evolution model in which tumor cells proliferate until an advantageous genotype is achieved, whereupon the clone will undergo clonal expansion and form the mass of the tumor. Furthermore, the study confirms the knowledge that genomic profiles of metastases are very similar to their primary tumors and probably emerge from individual cells out of the tumor mass at a late stage in cancer development. After seeding, the metastases only undergo further mutations, but usually don't develop new chromosomal breakpoints. While the SNS method succeeds in profiling single cells, the thin sequence coverage makes this analysis problematic. As an alternative, a combination of multiple cells from a defined subpopulation could be used to enhance regional and global genome coverage. In addition, since all SNS studies only analyze a limited number of single cells, compared to the billion of cells a tumor or metastasis has, the observed picture might be missing key components. The most lethal tumorigenic subpopulation must not always be the most prominent fraction and thus, it is possible that these SNS studies are not be able to identify such populations.

Barcoding

In 2013, Nolan-Stevaux utilized DNA barcoding technology to study tumor heterogeneity (Nolan-Stevaux et al. 2013). For this, they used lentiviral barcoding to measure the growth heterogeneity of cancer cells in xenografts in immune-deficient mice. They infected colorectal cancer cell lines with a lentiviral barcode library, containing unique nucleotide barcodes. Barcoded cells were injected into mice, xenograft tumors were resected and genomic DNA was extracted. The barcode was then identified using NGS. Analysis of

the *in vivo* distribution revealed that 75% of the injected cell clones did hardly engraft in mice and only made up 1% of the overall tumor mass. On the other hand, 7% of the injected tumor clones were able to divide at an above-average rate, leading to around 80% of the total number of tagged cells in the tumor. The phenomenon that a small number of individual cancer cell clones from a heterogeneous pool takes over the entire tumor after implantation is termed clonal dominance. As an explanation, the authors presumed the interplay of two stochastic contributing factors. For one, each cell in a homogeneous cancer population exhibits a slightly different phenotypic state, which ultimately converge to equilibrium (Gupta et al. 2011). Second, since the tumor poses a highly diverse environment for cells with varying regions of hypoxia and necrosis, clones in a favorable area of the tumor (e.g. close to blood vessels), exhibit a positive selection compared to clones in a less favorable geographic area. The authors came to the conclusion that clonal dominance is driven rather by randomness than by genetic differences, referring to a study in transplanted xenografts in which cells showed a statistically oscillating behavior (Kreso et al. 2013). Overall, Nolan-Stevaux employed a new and promising method for tracking different cancer cell clones *in vivo*, which could have a number of different applications in therapeutic research, for example to study the susceptibility and resistance of different clones during cancer treatments. One of the major drawbacks of this study is that they only used 3 animals for their study. It is possible that the reason why the authors did not detect an overlap of the dominant clones between the replicates and thus concluded it is driven by randomness is due to the limited amount of mice they used in their study.

The method of barcoding is a promising tool in characterizing tumor subpopulations and studying the development of tumor heterogeneity. We decided to employ the barcode strategy on the 4T1 murine breast cancer cell line in order to identify and track individual

clones throughout tumor progression and metastasis. Combining this approach with experimental cancer therapies can shed light on possible new targets and improved cancer treatment strategies.

Clinical impact of tumor heterogeneity

Intra-tumor heterogeneity has large implications for the treatment of cancer patients. Currently, cancer diagnosis heavily relies on biopsies of a very small region of a tumor during diagnosis. However, the concept of tumor heterogeneity renders this method almost impossible. Presently, pathologists examine multiple sections from the same tumor, but report only the highest grade. This highly underestimates the somatic mutational landscape of a tumor. The assumption that tumor characteristics are passed on to metastases during disease progression has not always been proved correct, as seen in the studies presented above. This shortcoming is one of the major reasons, why cancer diseases are able to adapt to a changing microenvironment and eventually resist advanced drug therapy (Keats et al. 2012; Schuh et al. 2012).

Currently, targeted therapy usually tries to exploit a tumor's dependency on critical proliferation or survival pathways. This has led to significant prolonged survival rates and overall benefit for cancer patients. This approach can be effective, when clonal cell lines cooperate to enhance tumor growth (Snuderl et al. 2011; Bach et al. 2001). Here, several subclones occupy different niches within the microenvironment, which increases the survival of the entire tumor structure. Relationships between single subpopulations are complex and can vary, but each clone influences the growth of the whole tumor. Targeting one subpopulation can thereby weaken the entire tumor complex and can lead to temporary improvement of tumor burden. However, most studies come to the conclusion that tumor

heterogeneity leads to overall decreased sensitivity to targeted therapeutics (Szerlip et al. 2012). As different tumor subclones can differ greatly in their mutation and cellular signaling pathways, often only few clones can be targeted efficiently. As other remaining clones usually take over quickly, the clinical benefits are rather short-lived (Saunders et al. 2012).

In order to develop therapies targeting a broader range of subclones, driver mutations need to be identified, which occur early during tumor development. As these mutations are ubiquitous, therapeutic approaches targeting these mutations have the potential to be much more effective (Saunders et al. 2012; Swanton 2012). However, even targeting these driver events might not have a significant effect on therapy outcome, since small tumor subpopulations continue to evolve and could develop drug resistance. This has been observed previously for mutated EGFR in non-small cell lung cancer (Su et al. 2012; Lynch et al. 2004). Thereby, a later mutation was able to abrogate the tumor's dependency on an earlier occurring driver mutation. Thus, it is likely that therapeutic success and patient survival is largely dependent on the frequency of somatic mutations along with the number of subpopulations in the tumor.

Tumors developing drug resistances are indeed the biggest problem faced currently by oncologists. Once metastatic solid tumors have formed, almost all cancer patients die of their disease. Usually drug resistance is conferred by a small subpopulation harboring one specific genomic mutation that can evade first line treatment, as it was seen in a study on AML (Ding et al. 2012).

Although it is largely known that conventional therapies are oftentimes not able to cure cancer, they are used to prolong survival times of many patients. Indeed, in some cases tumors respond very well to treatment. However, in worst cases chemotherapy even

increases the chances of relapse by inducing selective evolution of tumor clones. It was observered that interval cytotoxic chemotherapy increases subclonal evolution, even contributing to relapse occurrences (Ding et al. 2012). This is not surprising, considering that chemotherapy causes damaging of the DNA and increases overall genomic instability.

While in some cases drug resistance is developed over time, these mutations can even exist before treatment and impart significant selective advantage of one population. After targeted therapy kills previous dominant clones, this population can proliferate and take over the tumor. It is known that some patients with chronic myeloid leukemia possess an Imatinib-resistance in the BCR-ABL fusion gene in some subclones (Shah et al. 2002). Thus, while the standard used Imatinib therapy kills most of the leukemic cells, a resistant minority can proliferate and ultimately lead to the death of the patient. Knowledge of every subpopulation in a tumor would enable the application of individual combinatorial approaches, in order to target all subclones at once. This is most likely achieved with a combination of several cancer drugs, delivering an all-out attack.

The viral barcoding strategy is a promising approach to study the effects of tumor growth, metastasis, drug resistance and microenvironment on different tumor populations in mouse models. We combined the barcode strategy with single cell sorting of individual clones so that each clone can easily be isolated and tested functionally. This knowledge can then be translated into potential clinical approaches.

Chapter 2: Transcriptional framework of ductal carcinoma in situ (DCIS)

Ductal carcinoma in situ (DCIS) is a spectrum of abnormalities that initiate in the cells that line the breast ducts (Barnes et al. 2012; Virnig et al. 2010). DCIS is considered a noninvasive form of breast cancer because the abnormal cells are still confined to the original layer of cells. DCIS may or may not progress to invasive ductal carcinoma (IDC). We wanted to combine laser capture microscopy and low input RNAseq, in order to profile human patient samples diagnosed as having DCIS or DCIS with micro-invasion. We were able to process DCIS, IDC, normal ducts and stroma (any surrounding connective tissue) from multiple patient samples and generated high-quality RNAseq libraries. Hierarchical clustering of these samples revealed that all of the DCIS samples clustered tightly together, whereas the IDC samples clustered in multiple subgroups. We were able to identify subtypes of DCIS and IDCs based on existing intrinsic breast cancer markers. It is our hope that this will lead to predictors for progression of DCIS to IDC. Lastly, we carried out differential expression between DCIS vs. normal ducts and DCIS vs. IDC samples in order to identify genes implicated in disease progression and invasion. Some of these differentially expressed genes have been shown to be involved in metastasis and we are hoping that this will lead to the identification of new biomarkers.

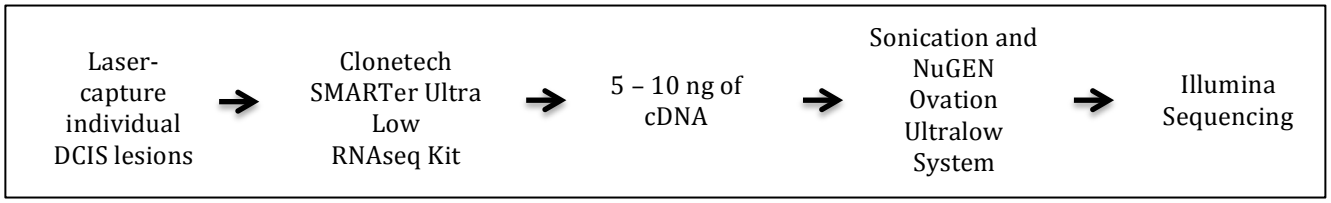
Transcriptional profiling of DCIS

We wanted to evaluate the biology of human DCIS through transcriptome sequencing (RNAseq). We used next-generation sequencing to generate gene expression profiles from single DCIS lesions of human patient samples. We wanted to know whether DCIS lesions could be classified to subgroups, similar to the 'intrinsic' subtypes of breast cancer. In addition, we wanted to compare individual DCIS lesions to IDC regions and normal tissue from the same patient in order to identify genes that might be responsible for the invasive phenotype.

We have established a collaboration with the Duke SPORE breast tumor bank (Dr. H. Kim Lyerly and Dr. Joseph Geradts), which contains 7,000 core needle samples from 1,700 patients, including 222 cases pure DCIS tumors. These samples are pathologically confirmed and annotated with HER2, ER, PR, and lymph node status. The full clinical annotation of these specimens is currently in progress.

We generated large-scale transcriptome datasets that describes the molecular events that occur in DCIS and invasive carcinoma (Fig 2.1a,b). Overall 10 DCIS tumors were processed for transcriptome sequencing. For each DCIS tumor, we isolated multiple distinctive DCIS lesions, invasive carcinoma, stromal cells and normal tissue using the Leica LMD 7000 laser capture microscope and processed them for RNAseq. The histology of each DCIS lesion was documented and post-analyzed by the pathologist, Dr. Joseph Geradts, at the Duke SPORE breast tumor bank.

a



b

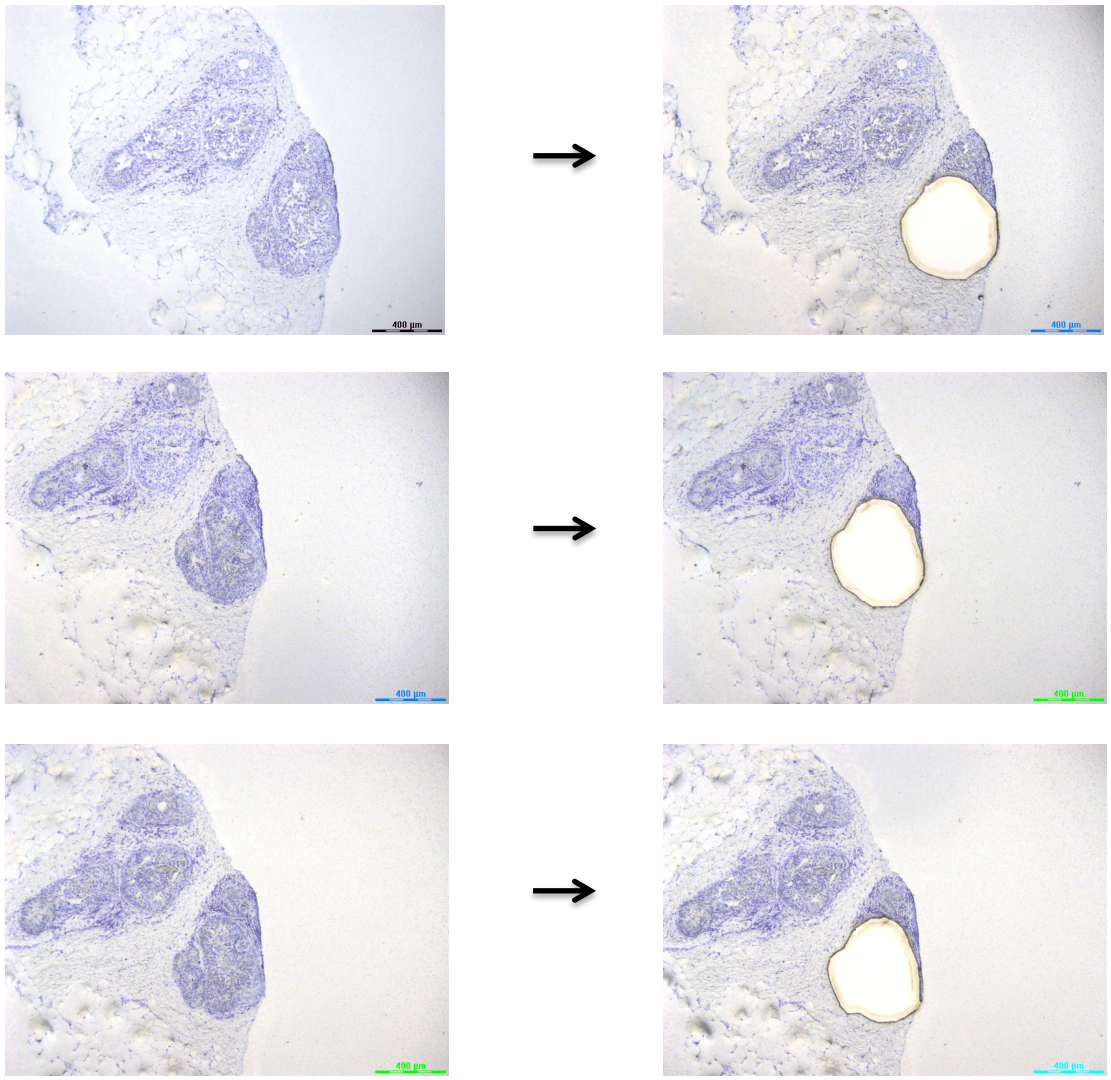


Figure 2.1 Workflow of DCIS isolation and RNAseq library construction
(a) Schematic of RNAseq library construction. The Clontech SMARTer Ultra Low technology uses oligo(dT) priming to generate cDNA from as low as one single cell. After sonication of the cDNA, the NuGEN Ovation Ultralow System is used to generate Illumina sequencing libraries. (b) Examples of individual DCIS lesions being cut by the laser capture microscope. For each sample, three serial sections are individually cut and processed to account for biological replicates.

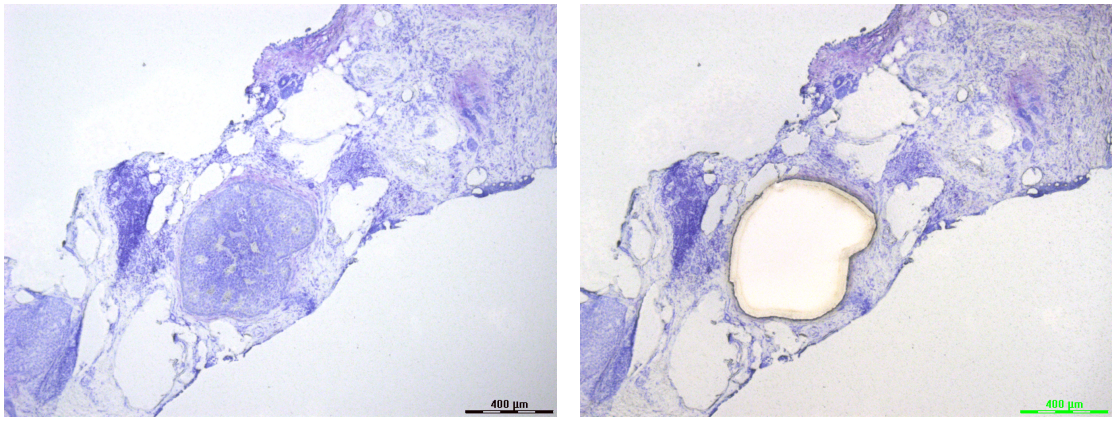
Clustering of transcriptome data

From the 10 breast tumor patient samples we processed, we were able to generate 117 RNAseq libraries that passed quality control (high percentage of unique mappers, Table 2.1). Five of these patient samples contained only pure DCIS and the other set of five samples contained a mix of DCIS and IDC. Overall, the samples were derived from multiple DCIS and IDC lesions, stroma and normal tissue (Fig 2.2). Each of these samples were prepared from roughly 50 – 100 cells. After mapping and gene abundance estimation, all genes showing at least 10 reads per million per nucleotide (RPM) were used to cluster the transcriptome data (Fig 2.3). The most notable cluster is that encompassing nearly all DCIS samples in the dataset (labeled with blue bars below the dendogram). Within the DCIS cluster, the profiles cluster by patient origin (denoted by the color of the sample name). For each sample, there are two to three biological replicates, which can be identified by the same sample name. The biological replicates almost always cluster together. The other notable clusters in Figure 2.3 are those that comprise most of the normal and stromal tissues (denoted by black and green bars, respectively). Most of the IDC samples (denoted by red bars) clustered away from the DCIS samples and were closer to the normal and stromal tissues. There was a smaller cluster of IDC samples to left of the DCIS cluster, which we denoted IDC-A and a bigger cluster of IDC samples to the right of the DCIS and stroma cluster, which we denoted IDC-B. There could be several explanations for these multiple IDC subgroups. There might be multiple molecular subtypes for IDC, whereas all DCIS lesions might be more similar to each other. Because we only had a limited number of IDC samples, we might not able to the see the full spectrum of different subgroups with the IDC class. Another explanation might be a technical limitation for isolating IDC lesions. Because the cells have broken off from a relatively pure DCIS lesion, they are mixed with other

Table 2.1: Overview of the number of samples collected, processed and analyzed

Number of patient samples dissected	10
Number of RNAseq libraries constructed	141
Number of RNAseq libraries that passed quality control	117

a



b

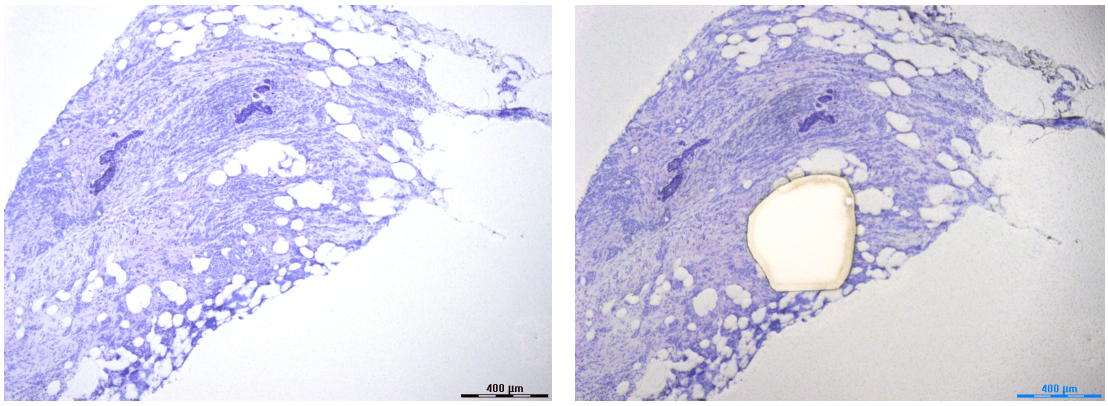


Figure 2.2 Histological comparison between DCIS and IDC
(a) Pre- and post-cut of a DCIS lesion (b) Pre- and post-cut of an IDC lesion. Note the heterogenous cell population within the IDC lesion.

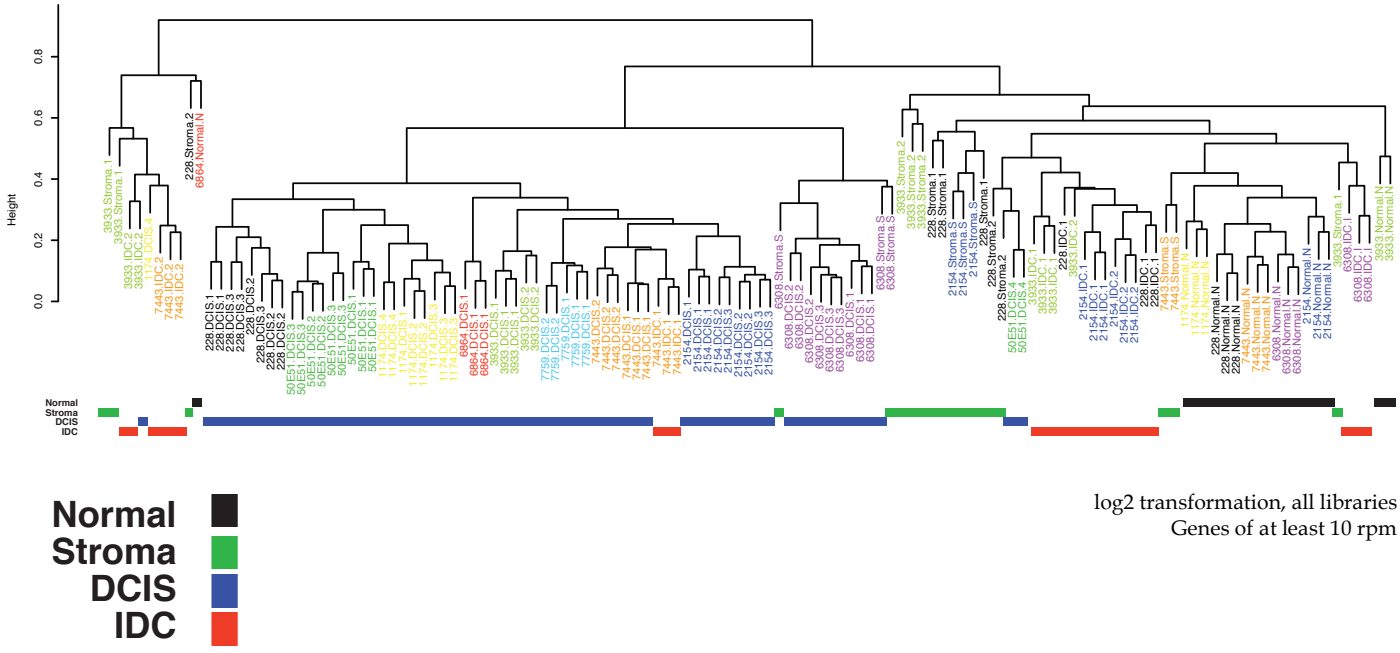


Figure 2.3 Sample clustering based on transcriptome data
RNAseq libraries were clustered based on their respective gene expression (only genes of at least 10 rpm). Sample types (DCIS, IDC, normal ducts, stroma (any surrounding connective tissue cells)) are color-coded. For each sample, there are two to three biological replicates, which are denoted by the same sample name. Sample names are color-coded based on the patient origin. Most DCIS samples cluster together, whereas stroma, normal and IDC samples are broken into several clusters.

stromal and normal cells (Fig 2.2). Thus, there might be some stromal contamination associated with the IDC samples and this might be the reason, why they cluster closer to normal and stromal samples. For future analysis, it will be important to quantify the amount of stromal contamination from each IDC lesion and discard the samples that possess a high stromal infiltration.

In addition, the sample type (DCIS, IDC, normal, stroma) can be separated by a principal component analysis (Fig 2.4a). If we look at the patient origin of the same principal component analysis, we can see that there is no clear separation based on patient ID (Fig 2.4b). Thus, we can exclude large batch effects on these RNAseq libraries. However, there are small patient-to-patient variations and it will be important to account for these. We are collaborating with Prof. Molly Hammell to use a generalized linear model (GLM) to improve the analysis on our differential expression (Auer and Doerge 2010). Basic differential analyses handles simple pair-wise comparisons, for example DCIS vs. IDC, but these analyses can't account for other variations. GLM performs multi-variable comparisons and calculates the variation contributed by each variable (for example tumor type or patient). We can then determine the major contributor of variation on the gene expression alterations and then account for these effects.

Subtype classifications based on breast cancer markers

We wanted to see if we could identify any DCIS/IDC subtypes using existing 'intrinsic' breast cancer markers. For this we used PAM50, which uses a minimal gene set of 50 genes in order to classify subtypes of breast cancer (luminal A/B, Her2-enriched, basal and normal-like subtypes) (Parker et al. 2009; Tibshirani et al. 2002). This gene set has a high agreement with other larger gene sets previously employed for subtyping and is now

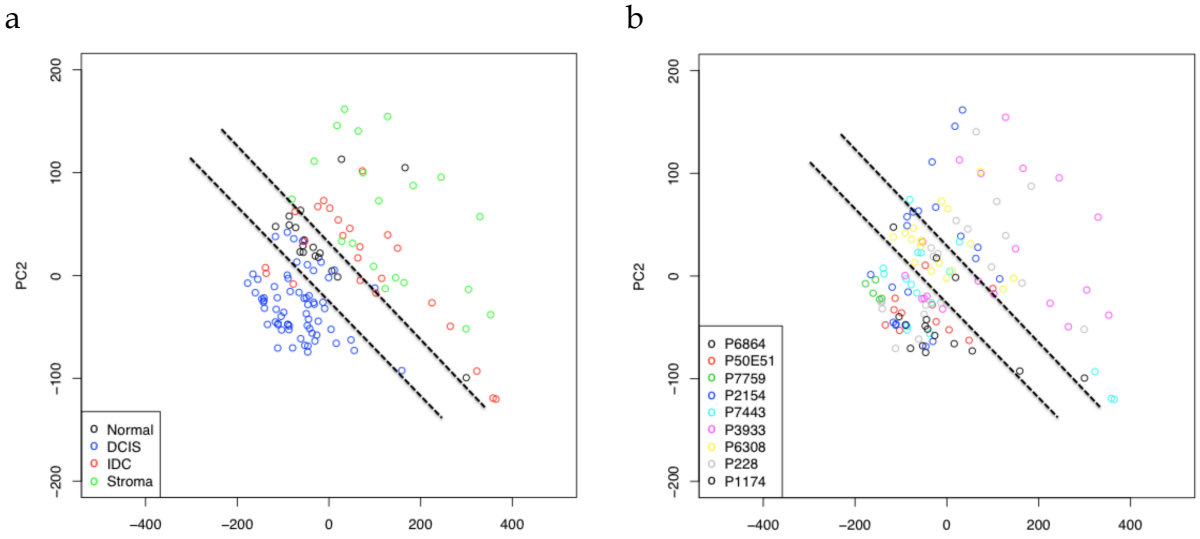


Figure 2.4 Principal component analysis on transcriptome data
(a) Principal component analysis separates samples based on type. (b) Principal component analysis separates poorly based on patient ID. There seems to be no large batch effects.

commonly used (Esserman et al. 2012; Gonzalez-Angulo et al. 2012; Perou et al. 2000). We were able to cluster the samples into several subtypes (Fig 2.5). Sample names are color-coded (blue for DCIS, red for IDC and black for normal). We were able to identify a cluster with Her2-enriched DCIS with high expression of HER2. A second cluster was represented by ER/PR-positive DCIS and IDC with high expression of estrogen receptor 1 (ESR1/ER) and relatively high expression of progesterone receptor (PGR/PR). In addition, a normal/normal-like cluster with samples showing strong expression of cytokeratins (CK5, 14, and 17) and a basal-like IDC cluster showing hallmarks such as low expression of NDC80, ANILIN, MELK and CEP55 were identified. Lastly, there is a mixed DCIS and IDC group that is characterized by strong expression of SFRP1.

In summary we were able to generate subtypes of DCIS and IDC based on existing breast cancer markers. The DCIS samples seemed to fall either to the HER2-enriched cluster or to the ER/PR-positive cluster. The IDC samples fell into more clusters than the DCIS samples. However, a large proportion of the IDC samples fell into the basal-like cluster. Almost all of the normal samples were clustered into a single subtype. It will be interesting to correlate these subtypes with clinical annotations of our samples. The hope is that we will be able to predict clinical outcome based on the subtype classifications of the DCIS and IDC samples.

Differential gene expression analysis

We wanted to look at differential expressed genes between DCIS vs. normal and DCIS vs. IDC in order to identify genes that might regulate the DCIS or invasive phenotype. Table 2.2 shows the top 20 up-regulated genes in all DCIS vs. normal (based on abundant genes above 100 RPM). Not surprisingly, ERBB2/HER2 is highly up-regulated in DCIS

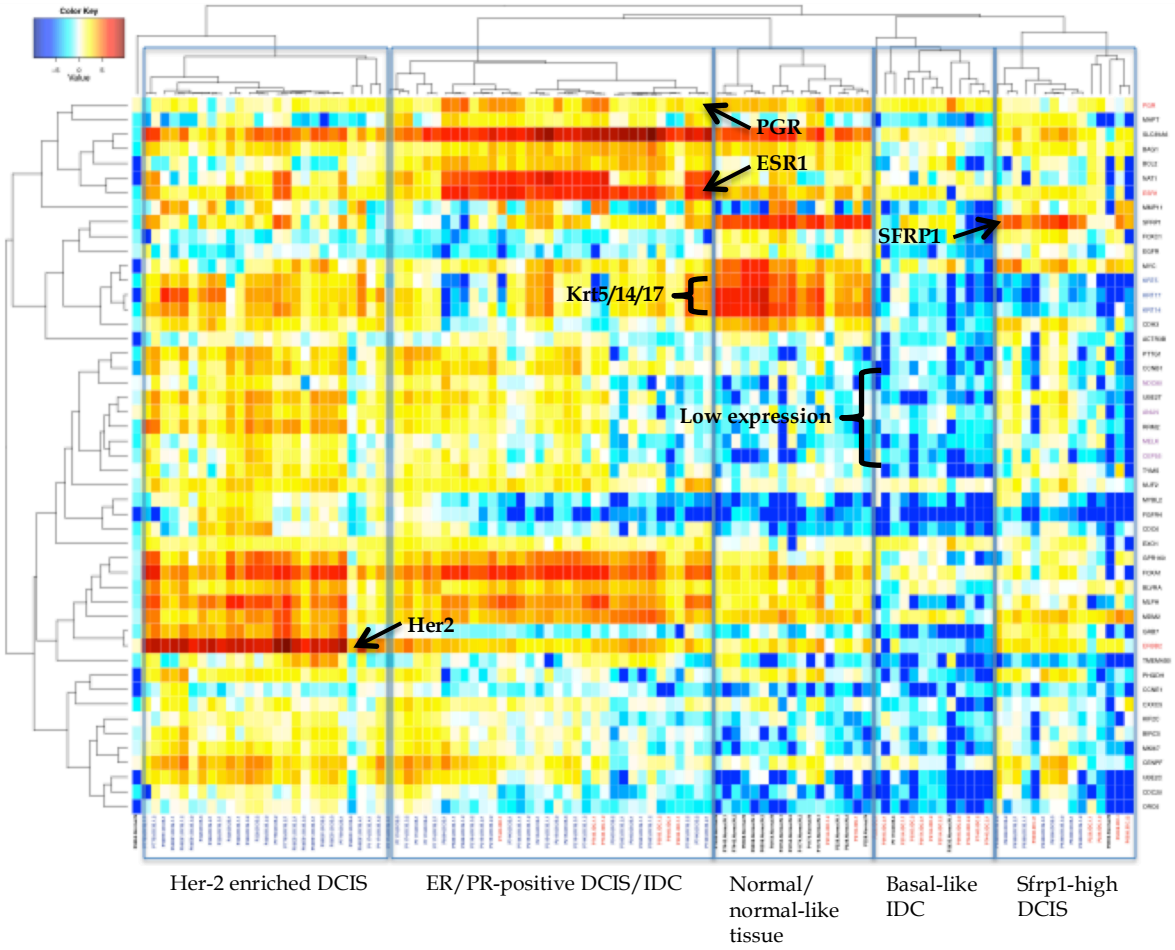


Figure 2.5 Subtype classifications of DCIS and IDC
The PAM50 gene list was employed in order to generate subtype classifications of all DICS, IDC and normal samples. DCIS samples are color coded in blue, IDC samples are in red and normal samples are color coded in black. Five clusters of samples were identified using this gene set.

Table 2.2: Top 20 up-regulated genes in DCIS vs. normal

Gene ID	Normal [counts]	DCIS [counts]	log2 fold change	p value
S100A7	3.292807589	2263.628097	9.425103022	7.88E-07
DCD	5.992251298	1353.854982	7.819759371	0.001692923
CLEC3A	4.828873426	467.7263374	6.59783421	1.88E-05
SLC30A8	8.467888827	700.853813	6.370967413	7.10E-14
CPB1	34.1629011	2264.310401	6.050497439	0.008263425
S100P	7.047804263	353.9492472	5.65022484	2.07E-07
ABCC11	8.666465162	330.6280151	5.25362149	0.002430132
GSDMB	5.091157587	162.9579835	5.000362501	0.00033062
PLA2G2A	3.569441084	113.8062982	4.9947384	0.00038475
SOWAHA	5.10104815	139.1098849	4.76928741	1.72E-06
S100A8	129.2339085	2932.441724	4.504045878	0.005038106
ERBB2	195.4103404	3261.090965	4.060775968	8.75E-07
THRSP	9.278249621	153.1166878	4.044635056	0.013033397
CST4	6.410298083	100.9750141	3.977463092	0.004450301
FADS2	12.47273338	174.7953615	3.808817332	0.005894755
GRB7	14.77495461	202.0266309	3.773319876	5.97E-06
CYP4Z1	18.07051107	241.8349511	3.74231355	0.015515287
SERHL2	10.53707953	137.358577	3.70440003	0.001306822
CYP2B7P1	17.14755243	211.9770043	3.627833194	0.00054054
SCD	186.1981021	2260.790772	3.60191721	3.52E-07

Based on abundant genes above 100 rpm

compared to normal tissue. The highest log2 fold change was observed for S100 Calcium Binding Protein A7 (S100A7). This gene has previously been identified as a molecular marker for DCIS (Porter et al. 2003; Emberley et al. 2004; Petersson et al. 2007). Next, we wanted to identify genes that were overexpressed in all DCIS vs. IDC. Because we identified two major clusters of IDC (Fig 2.3), we compared all DCIS to the IDC-A cluster and all DCIS to the IDC-B cluster. As a reminder, IDC-A represents the smaller cluster of IDC samples to left of the DCIS cluster and IDC-B represents the bigger cluster of IDC samples to the right of the DCIS and stroma cluster. Table 2.3 and 2.4 shows the top 20 up-regulated genes in all DCIS vs. IDC-A and all DCIS vs. IDC-B respectively (based on abundant genes above 100 RPM). Some of the overexpressed genes have been implicated with invasion or metastasis. For example, lymphotoxin beta (LTB) belongs to the TNF superfamily and has been associated with higher expression in lymph node metastasis compared to matched primary papillary thyroid carcinoma (Cerutti et al. 2007). E-selectin (SELE) was shown to be higher expressed in metastatic colorectal cancer cells compared to local non-metastatic variants (Aychek et al. 2008). Interestingly, both tables identified DARC and CCL19 as up-regulated genes. DARC is an atypical chemokine (Duffy Blood Group) and CCL19 is chemokine ligand (C-C Motif). Many publications support the notion that DARC and CCL19 are both expressed by stromal cells. In fact, DARC is a decoy cytokine receptor that is expressed on vascular endothelial cells and is known to trigger senescence in tumor cells (Bandyopadhyay et al. 2006; Iiizumi et al. 2007). The chemokine CCL19 is released by lymphatic endothelial cells and T-cells of the lymph node and act as ligands of CCR7. It is known that CCR7 expression is highly expressed in breast cancer cells and is related to lymph node metastasis (Nakata et al. 2008; Müller et al. 2001). Because some of the up-

Table 2.3: Top 20 up-regulated genes in DCIS vs. IDC (subgroup-A)

Gene ID	DCIS [counts]	IDC-A [counts]	log2 fold change	p value
CCL21	0.170349073	688.2937044	11.98031451	2.06E-16
LTB	0.847375119	180.4590634	7.73445512	1.45E-14
CCL19	8.651999547	1683.222988	7.603977008	1.18E-13
DARC	1.874469229	363.3545112	7.598751865	8.94E-50
PLA2G2D	0.602246968	109.0443122	7.500343579	1.17E-12
CD247	0.900710435	106.9000804	6.890983847	6.12E-20
CD2	3.399181133	395.1170182	6.860948935	5.41E-28
IL7R	8.225440724	922.8725355	6.809894609	2.35E-50
MS4A1	4.190367432	398.2318704	6.570388122	3.03E-05
CD3E	1.157367137	107.2629222	6.534161067	1.13E-14
GZMK	2.347138292	207.6398944	6.467037002	1.87E-20
LCK	2.186226579	193.2745088	6.466064632	1.72E-21
CSF2RB	4.625420386	362.1313371	6.290784723	1.13E-15
CCR7	1.858701	144.4863406	6.280494591	1.40E-16
CD52	16.92475936	1260.471406	6.218684259	1.81E-15
CD48	6.992472573	516.6267999	6.207175985	1.76E-25
IL2RG	4.495106588	324.9917898	6.175904136	6.68E-21
CD3D	5.167625796	366.242134	6.14715236	4.00E-11
PTPRCAP	2.276762187	147.2180352	6.014827005	1.44E-13
PTGDS	5.431382056	342.3696934	5.97809184	1.25E-19

Based on abundant genes above 100 rpm

Table 2.4: Top 20 up-regulated genes in DCIS vs. IDC (subgroup-B)

Gene ID	DCIS [counts]	IDC-B [counts]	log2 fold change	p value
DARC	1.871324305	159.4122348	6.412558947	2.01E-05
SELE	3.865334705	205.1276593	5.729784883	0.000674387
DSG1	5.183437809	182.7749067	5.140014952	5.13E-11
IGLL5	41.20435265	1367.589955	5.052695176	0.007256623
LOC646736	4.864919275	141.8743084	4.866053684	3.62E-09
SNORA70F	5.980372354	171.3459661	4.840533104	1.12E-08
LINC00864	4.253391352	114.9093593	4.755738897	7.28E-12
PLSCR5	7.90720515	203.5293182	4.685924963	4.97E-08
LOC100144595	12.14756489	309.7768164	4.672490133	1.52E-05
UGT2B4	8.358162067	212.36979	4.667249012	0.000593047
PMCHL2	11.76286395	296.0673512	4.653614139	5.53E-21
CCL19	8.439768349	208.5736633	4.627209789	0.000270716
LAMB4	14.68078336	357.314458	4.605193434	5.40E-10
UTS2B	4.464203884	108.1220261	4.598113724	0.000242235
PGAM1P5	6.813545769	163.0010807	4.580331947	0.00036807
CDH8	6.08313759	145.0577473	4.575667904	0.000163982
C9orf153	4.773991355	112.6460356	4.560456776	2.82E-06
LOC152578	4.981497753	117.3977404	4.558681258	4.58E-15
KLHL1	4.954127604	116.424215	4.554616316	3.41E-05
NXPH2	10.76034935	248.7185469	4.53071727	4.83E-05

Based on abundant genes above 100 rpm

regulated genes in IDC seem to be preferentially expressed in stromal cells, there might be some stromal contamination in the IDC lesions that we collected.

In summary, we were able to identify differentially expressed genes that have previously been to shown to be involved in DCIS or IDC. In the future, it will be important to exclude IDC samples with high stromal contamination in order to avoid false-positive in the differentially expressed gene list. In addition, we are planning to analyze the gene expression signature from the surrounding stroma, for example to identify differentially expressed genes between stroma that is adjacent to DCIS vs. stroma that is adjacent to IDC. One hypothesis is that the stroma might be important for determining whether disease actually progresses to invasive carcinoma. Overall, it is our hope that we will be able to identify biomarkers that might have predictive power over patient outcome.

Methods

Tissue collection

Core needle samples from the Duke SPORE breast tumor bank were shipped on dry ice to Cold Spring Harbor Laboratory for processing. These tissue samples were cut into multiple 10 μm thick sections, mounted to RNAse free PEN membrane slides (Leica) and quickly stained with Toluidine blue (Sigma). For each DCIS patient sample, we isolated multiple distinctive DCIS lesions, invasive carcinoma, stromal cells and normal tissue using the LMD 7000 laser capture microscope (Leica). Three serial sections were cut from each distinctive tissue lesion in order to provide biological replicates for the RNAseq library generation. Tissue sections were collected in Prelude Direct lysis buffer (NuGEN). Images of each tissue lesion were taken pre- and post-dissection. The histology of each tissue lesion was documented and post-analyzed by the pathologist, Dr. Joseph Geradts, at the Duke SPORE breast tumor bank.

RNA-Seq library preparation

Total RNA from dissected individual tissue lesions were converted to double-stranded cDNA using the SMARTer Ultra Low Input HV system (Clontech). cDNA was fragmented using the Covaris LE220 sonicator according to the manufacturer's instruction to yield a target fragment size of 200 bp. About 10 ng of amplified cDNA was used to construct Illumina sequencing libraries using the Ovation Ultralow Library System (NuGEN). Sequencing was performed on the Illumina HiSeq 2000 instruments.

Analyses

RNAseq samples were mapped to the human genome (hg19), and unique reads were annotated against RefSeq (Release 59) gene model as described above. Gene abundances were obtained using HTSeq package, and normalized by millions of mapped reads in each library (Reads Per Million, RPM). Genes with less than 10 RPM were removed and for the rest, expression values log2-transformed. Hierarchical clustering was then performed based on Pearson correlation values, with cluster distances computed using complete linkage.

For the Principal component analysis, normalized gene abundances were obtained from RNAseq samples as previously described, and log2-transformed. PCA was performed, and the samples were plotted along the first two principal components (PC1 and PC2). Samples were color-coded either by their tissue type (DCIS, normal, IDC or stroma), or the patient ID.

For the subtype classification, normalized gene abundances were obtained from RNAseq samples as previously described. A heatmap of genes belonging to the PAM50 set of breast cancer subtype markers was generated from log2-transformed gene abundances, with hierarchical clustering of both samples (DCIS, IDC and normal) and PAM50 genes using Pearson correlation values and complete linkage for cluster distances.

Chapter contributions

Sara Gutiérrez-Ángel assisted with laser-capture microscopy and RNAseq library preparation, Dr. Oliver Tam and Prof. Molly Hammel conducted bioinformatics analysis and Dr. Joseph Geradts assessed histology of post-cut specimens.

Chapter 3: Functional genomics of disseminated and circulating tumor cells in a breast cancer mouse model

Breast cancer starts as a local disease, but can spread to distant organs through a multi-step process. Malignant cells from the primary tumor enter the blood circulation and ultimately metastasize to distant organ sites, such as lung, liver and bone. We have utilized a well-established mouse model of breast cancer metastasis to study individual aspects of the metastatic cascade. We were able to show that a non-metastatic mouse breast cancer cell line was able to metastasize to the lung upon T-cell immune suppression. Exome capture sequencing was carried out in order to identify potential neo-antigens that might be recognized by these T-cells.

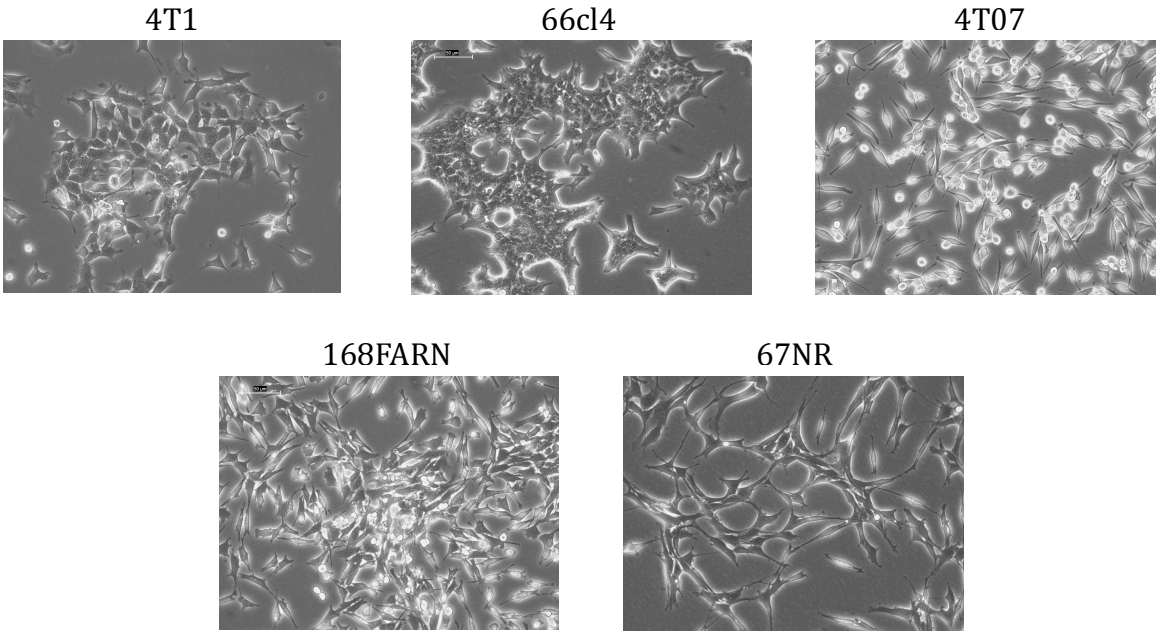
In addition, we have used these mouse breast cancer cell lines to look at the biology of circulating tumor cells (CTCs) in the blood. For this, we fluorescently tagged tumor cells in order to isolate CTCs, metastatic tumor cells in the lung and liver and primary tumor cells for gene expression profiling. The goal of this approach was to identify genes that contribute towards metastasis formation. We decided to focus on genes that were overexpressed in CTCs vs. the primary tumor. From this gene list, we were able to identify and functionally characterize Ptpn22 as a metastatic inducer using a pooled *in vivo* shRNA screen. Knock-down studies using shRNAs against Ptpn22 revealed a reduction in overall lung metastasis. This is due to a decrease in the number of CTCs in the blood circulation upon Ptpn22 knock-down. *In vitro* assays that mimic the growth and survival of CTCs in the blood circulation revealed a higher apoptotic rate upon reduction of Ptpn22 levels. Taken together, we were able to identify a novel metastatic regulator that specifically acts on the survival of CTCs.

Metastatic potential of the 4T1 cell line series

The 4T1 mouse mammary tumor model consists of five distinct tumor cell lines, called 4T1, 66cl4, 168FARN, 4T07 and 67NR. All of these lines have been derived from a single spontaneous mammary tumor in a wild-type BALB/c mouse (Dexter et al. 1978; Aslakson and Miller 1992; Miller et al. 1983b). The 4T1, 66cl4 and 168FARN cells display an epithelial morphology, whereas the 4T07 and 67NR show a more mesenchymal morphology (Fig 3.1a). When these cells are orthotopically injected into the mammary gland of syngeneic BALB/c mice, they form mammary tumors with similar kinetics. However, their metastatic potential differ significantly (Fig 3.1b) (Aslakson and Miller 1992). The 4T1 cell line is the most aggressive cell line and readily metastasizes to distant organ sites including lung, liver, bone and brain within 3 – 4 weeks. The 66cl4 cell line is weakly metastatic and primarily metastasizes to the lung through the lymphatic system (Miller et al. 1983a). The 4T07 cell line is able to disseminate from the primary tumor, however is not able to colonize any distant organ sites (Miller et al. 1987). The 168FARN cell line metastasizes to the adjacent lymph node and the 67NR cell line is non-metastatic. For both 168FARN and 67NR, no tumor cells are detectable at distant organ sites (Miller et al. 1986). Interestingly, the 4T1 and 4T07 cell lines are derived from the same parental cell line called 410.4 (Aslakson and Miller 1992). The 4T1 cell line is a thioguanine-resistant variant that was selected from the 410.4 cell line without mutagen treatment. On the other hand, the 4T07 cell line is a thioguanine-resistant and ouabain-resistant variant that was treated with mutagen (Miller et al. 1987).

We wanted to validate the metastatic potential of these five different cell lines in our hands. Thus, we injected 2 x 10^6 cells of each cell line separately into the mammary gland of syngeneic BALB/c female mice. After 4 weeks, we harvested the primary tumors, lungs and liver and processed them for Hematoxylin and Eosin (H&E) staining. We were able to detect

a



b

	67NR	168FARN	4T07	66cl4	4T1
Metastasize to:					
Lymph node	–	+	+	++	+
Blood	–	–	+	+/-	++
Lungs	–	–	+	++	+++
Liver	–	–	–	+	+++

Metastatic potential

Figure 3.1 Morphology and metastatic potential of 4T1 variants
(a) Phase-contrast images of the 4T1 variants *in vitro*. The 4T1, 66cl4 and 168FARN cell lines display an epithelial morphology, whereas the 4T07 and 67NR show a more mesenchymal morphology. (b) Schematic overview of the differential metastatic potential of the 4T1 variants. 200,000 cells were injected orthotopically into the mammary gland of BALB/c mice and the mice were sacrificed after 4 weeks (5 mice per cell line). For lungs and liver, the number of maco-metastases was evaluated based on H&E staining. Lymph node metastasis was evaluated based on H&E staining of the brachial lymph node. Blood was drawn and cultured *in vitro* in order to count the number of surviving colonies. 4T1 and 66cl4 cells are able to metastasize and colonize distant sites. 4T07 cells are able to disseminate from the primary tumor, but are unable to form distant metastases. 168FARN metastasizes to the local lymph node. 67NR is non-metastatic.

micro- and macro-metastases in the lung and liver of animals that were injected with the 4T1 cell line (Fig 3.2). The 66cl4 mice contained overt lung metastases, but no metastases were detected in the liver (Fig 3.2). The 4T07, 168FARN and 67NR did not display any metastases in the lung or liver based on H&E stainings. However, the 4T07 liver showed extensive sites of hematopoiesis (Fig 3.2), implicating an activation of the immune system upon injection of the 4T07 cells.

It is interesting to note that although 4T07 and 67NR show a more mesechnymal phenotype *in vitro*, which is often correlated to an epithealial-mesenchymal transition (EMT), they did not display an metastastic phenotype, when injected into syngeneic BALB/c animals. EMT is a process in which epithelial cells loose cell polarity and cell-cell adhesion to gain migratory and invasive properties and it is thought that this is a prerequisite for the initiation of metastasis (Tsai and Yang 2013). Thus, an EMT phenotype alone might not be sufficient for metastasis. However, there is also the possibility that other factors such as the immune system might suppress the metastatic phenotype of these cells (Bidwell et al. 2012).

4T07 cells are metastatic upon T-cell immune suppression

4T07 cells have no metastatic potential in wild-type BALB/c mice when injected into the mammary gland (Aslakson and Miller 1992). In contrast, we observed that injection of these cells into immunocompromised nude mice resulted in accelerated primary tumor growth that is similar to the primary tumor kinetics of 4T1 cells. Interestingly, we detected CTCs and lung metastases when the 4T07 cells are injected into nude mice, but not when they are injected into wild-type BALB/c animals (Table 3.1, see method section for isolation procedure). Nude mice can not develop mature T lymphocytes ($CD4^+$ and $CD8^+$) (Cordier and Haumont 1980). Since 4T07 cells are able to metastasize in an immunosuppressive

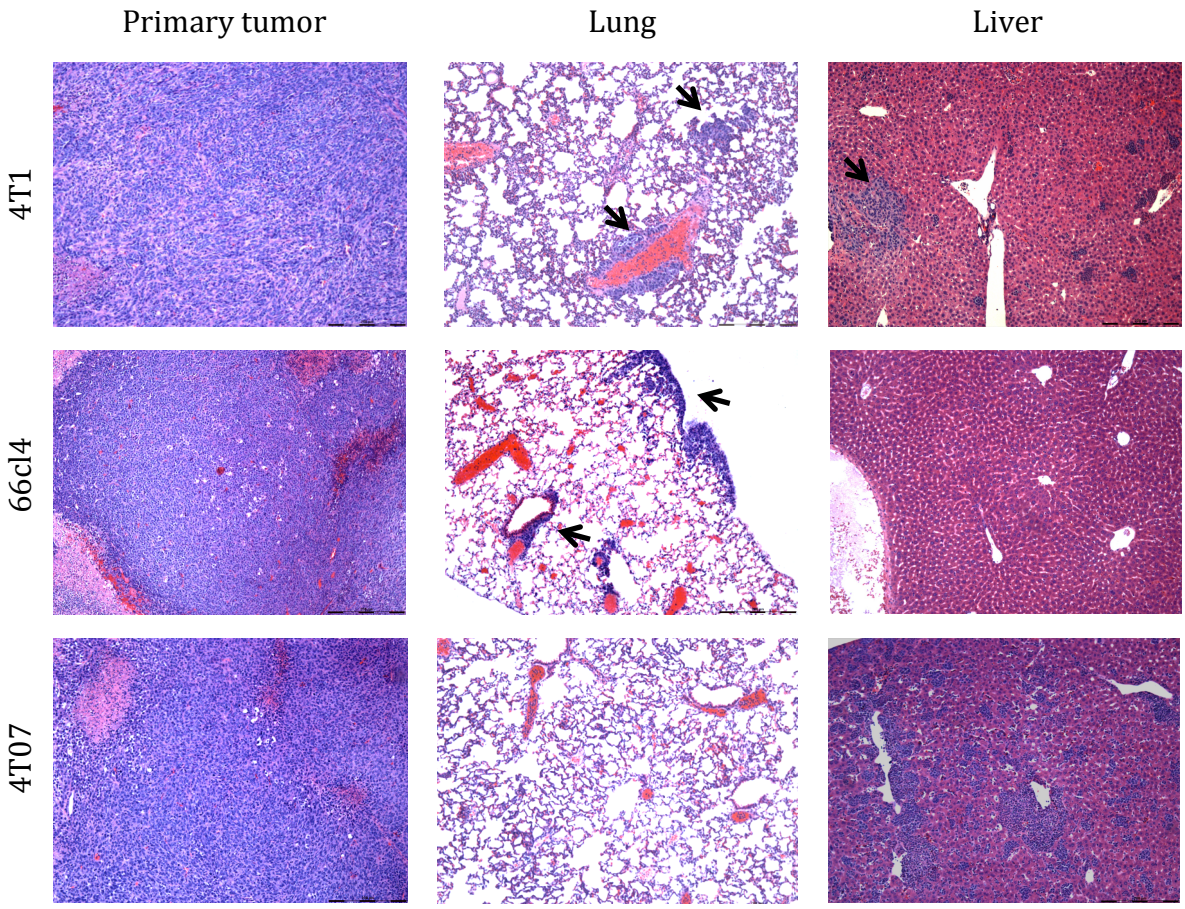


Figure 3.2 Histological analysis of primary tumor, lung and liver of 4T1, 66cl4 and 4T07 injected mice

2 x 10^6 cells were injected into the mammary gland of mice and primary tumor, lung and liver were harvested 4 weeks after injection. 4T1-injected animals displayed metastases in lung and liver, whereas 66cl4 injected animals showed only metastases in the lung. The 4T07-injected animals did not show any metastases in the lung or liver, however the 4T07 liver showed extensive areas of hematopoiesis. Arrows indicate metastatic areas. All primary tumors showed areas of necrosis.

Table 3.1: Metastatic profile of 4T07 cells orthotopically injected into Balbc/c or nude mice

Cell line	Lung mets [# of cells collected per mouse]	Liver mets [# of cells collected per mouse]	CTCs [# of cells collected per mouse]	Primary tumor growth [average time the mice reached endstage]
4T07 Venus **BALB/c**	0 cells	0 cells	0 cells	35 days
4T07 Venus **nude**	10,000 – 20,000 cells	0 cells	1000 – 2500 cells	22 days

n = 5 mice per group

environment, we hypothesize that these cells might be recognized and targeted by T-cells in an immunocompetent setting. In order to analyze this hypothesis we induced T-cell depletion in immunocompetent wild-type BALB/c mice by simultaneous anti-CD4 and -CD8 antibody treatment (intraperitoneal injection of 200 μg of each antibody every 5 days, Fig 3.3a). To check for proper T-cell depletion, the spleen of a mouse was isolated after a single anti-CD4 and -CD8 antibody injection, and no T-cells were detected by FACS analysis (Fig 3.3b). 3 weeks after injection of 4T07 cells and anti-CD4 and -CD8 antibody treatment, primary tumors and lungs were isolated. A large effect on primary tumor growth was observed between the two treatment groups. Animals with depleted $CD4^+$ and $CD8^+$ T-cells showed larger primary tumor growth than the control group, which was injected with PBS (Fig 3.3c). Analysis of lung metastases in H&E histological sections indicated an increased metastatic potential in T-cell depleted BALB/c mice in comparison to PBS treated control mice (Fig 3.3d). No lung metastases were found in any BALB/c control mice, even when the primary tumor reached the same size as the CD4/CD8 depleted mice. PBS was used as a control since abundant amounts of an isotype control antibody was not available.

In order to identify 4T07-specific T-antigens, we decided to carry out exome capture sequencing of the 4T1 cell lines (67NR, 168FARN, 4TO7, 66cl4 and 4T1). We generated exome capture libraries of all lines and identified single nucleotide variants (SNVs). The sequencing results revealed many SNVs between the different cell lines (Table 3.2). The cell lines 66cl4, 4T07 and 168FARN showed higher numbers of unique SNVs than the 4T1 and 67NR cell lines. This can be attributed to the fact the 4T1 and 67NR variants have not been treated with a mutagen, whereas the other variants were mutagenized during their selection process. We were interested in identifying SNVs that were present in the 4T07 cell line, but not in the 4T1 line in order to identify potential neo-antigens that might be recognized by T-

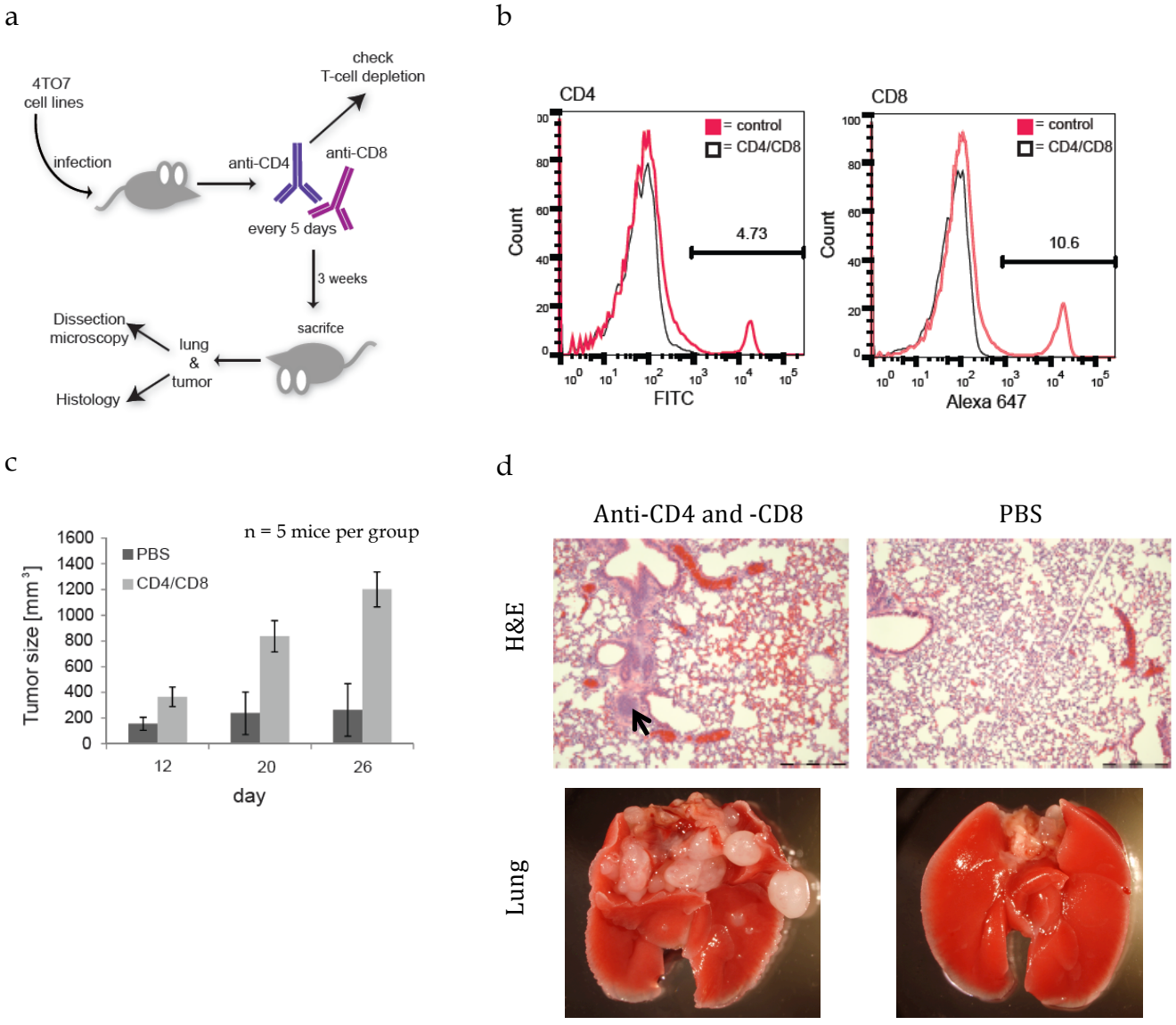


Figure 3.3 Depletion of CD4⁺ and CD8⁺ T-cells accelerates primary tumor growth and induces metastasis of 4T07 cells in BALB/c mice

(a) 4T07 cells are injected into the mammary gland of BALB/c mice and T-cell are depleted by simultaneous injection of anti-CD4 and -CD8 antibodies every 5 days. Animals are sacrificed after 3 weeks and primary tumor growth und lung metastases are evaluated by microscopy and histology. (a) FACS analysis shows depletion of T-cells in the spleen 4 days after a single injection of anti-CD4 and -CD8 antibodies. (c) Differential tumor growth between animals that were treated with anti-CD4 and -CD8 treatment and with PBS only. (d) Lungs of animals treated with anti-CD4 and -CD8 displayed large metastatic nodules under the microscope. H&E stainings revealed multiple micro and macro-metastases. Arrow indicates metastatic area.

Table 3.2: Exome capture sequencing of the 4T1 cell lines revealed many single nucleotide variants (SNVs)

Sample	Relative metastatic potential	EMS mutagenized	Total SNVs called	SNVs unique to this line
4T1	high	no	79229	65
4T07	low	yes	75350	781
66CL14	medium	yes	77105	601
168FARN	very low	yes	82357	746
67NR	none	no	111694	248

cells in the 4T07 system. For this, we overlapped these unique SNVs with our *in vitro* transcriptome data in order to identify SNVs in the 4T07 cell line that were highly expressed in the 4T07 variant vs. the 4T1 line (Table 3.3). One of the top hits was P2ry14, a G-protein coupled receptor. This receptor specifically responds to UDP-glucose and related sugar nucleotides (Abbracchio et al. 2003). In addition, there are algorithms available that will predict if antigenic peptides can bind to the major histocompatibility complex (MHC) class I molecules (Newell and Davis 2014). It will be interesting to see if any of the mutated peptides can be incorporated into the MHC groove for T-cell recognition.

In summary, we were able to detect accelerated primary tumor growth and overt lung metastases of 4T07 cells upon T-cell depletion in BALB/c mice. It would be advisable to repeat this experiment with an isotype control antibody, since there might be an increase in inflammation due to antibody treatment. The exome capture sequencing of 4T07 revealed many unique SNVs that are highly expressed in 4T07, compared to the 4T1 variant. We think that since both 4T07 and 4T1 cells were isolated from the same precursor variant, a direct comparison between both cell lines is useful. In order to see if these candidate gene products could display neo-antigens for T-cell recognition, a method called SEREX (serological analysis of recombinant tumor cDNA expression libraries) could be employed (Zhou et al. 2012). In this method, cDNA libraries are extracted from 4T07 primary tumors in order to immune-screen gene products recognized by T-cells. These clones are then directly sequenced and their identity can be determined. In addition, T-cell receptor sequencing might provide another alternative in order to identify endogenous pairs of T-cell receptor sequences from 4T07 injected mice.

Furthermore, in order to see whether 4T07 cells are able to circumvent the T-cell immune system in BALB/c mice by adaption, serial injections of 4T07 cells into BALB/c

mice could be performed. Doing RNAseq analysis of the cells after serial culturing and comparing them to the parental 4T07 cells, we might be able to narrow down potential T-cell antigens.

4T07 and 4T1 cells are derived from the same precursor cell line, whereas 4T07 cells were mutagenized with EMS, but not 4T1 cells. Thus, an EMS induced mutation in the 4T07 cell line might act as a neo-antigen for T-cell recognition. Given that this mutation was induced by EMS and not during tumor progression, this neoantigen might not be relevant for human breast cancer progression. However, the mechanism of how these neoantigen specific T-cells eliminate metastatic tumor cells might provide insight into designing effective treatments to induce immune responses against tumor cells.

Transcriptional profiling of primary tumor cells, CTCs and metastatic cells

We utilized the 4T1 mouse breast cancer cell line series, which consists of five distinct tumor cell lines (4T1, 66cl4, 4T07, 168FARN and 67NR). We generated gene expression profiles (RNAseq libraries) from these cell lines *in vitro* and also *in vivo*, by injecting them orthotopically into mice to isolate primary tumor cells, CTCs and metastatic tumors cells in lung and liver. We wanted to compare the gene expression profiles to each other in order to identify genes that are important for the metastatic process. We hypothesize that each of the defined metastatic potential displayed by these cell lines are partially due to alterations of gene expression in the various compartments (primary tumor, CTCs and lung/liver metastases).

To generate the gene expression profiles, we orthotopically injected fluorescently labeled ($Venus^+$) cells from each cell line separately into the mammary gland of nude mice. We decided to use immunocompromised mice since we detected engraftment issues with

Table 3.3: SNVs that were uniquely present and highly expressed in 4T07 (sorted based on log2 fold change expression 4T07/4T1)

Gene	SNVs: Cell Lines with Mutation	Expression: 4T07/4T1	log2(FC) 4T07/4T1	Q-value
P2ry14	4T07	145.03	7.18	1.75E-05
Col5a3	4T07	13.16	3.72	3.10E-14
Mtap1a	4T07	12.46	3.64	0.00E+00
Sema3b	4T07	7.49	2.90	0.00E+00
Sardh	4T07	6.33	2.66	1.44E-06
Thbs3	4T07,66CL14	5.34	2.42	2.73E-12
Adamts7	4T07	5.12	2.36	3.66E-09
Tmem98	4T07	4.27	2.09	1.62E-05
Herc3	4T07	3.67	1.87	1.88E-09
Tctn2	4T07	3.61	1.85	2.19E-09
Fyn	4T07	3.48	1.80	1.89E-10
Ank	4T07	2.83	1.50	1.01E-07
Sdf2l1	4T07	2.76	1.47	7.27E-04
Pdgfra	4T07	2.69	1.43	4.91E-11
Tmem39a	4T07,168FARI	2.35	1.23	0.00E+00
Eif2ak3	4T07	2.30	1.20	8.93E-04
Fzd2	4T07	2.17	1.12	2.26E-02
Zfp362	4T07	2.14	1.10	3.90E-04
Nedd4l	4T07	2.13	1.09	4.74E-15
Ipo4	4T07	2.08	1.06	3.90E-05
Adam15	4T07	2.08	1.06	0.00E+00
Unc93b1	4T07	2.01	1.00	2.28E-03

the 4T1 cell line series that carried various fluorophores such as tdTomato and Venus. The primary tumor growth rate and the metastatic potential were greatly reduced when these foreign antigens were incorporated into these cell lines and injected into immunocompetent BALB/c mice. At 4 weeks post-injection, primary tumor cells, CTCs from the blood and metastatic tumor cells from the lung and liver were isolated for gene expression profiling.

In order to isolate pure populations of tumor cells from the primary site, lung and liver, we enzymatically digested these tissues into single cells with collagenase and hyaluronidase and then sorted them based on fluorescence using FACS ($Venus^+$ cells, Fig. 3.4b). Similarly, we have developed a sensitive approach to isolate CTCs from the peripheral blood (Fig 3.4a). After whole blood collection by terminal cardiac puncture, erythrocytes were removed by red blood cell lysis and then $CD45^+$ leukocytes and $Ter119^+$ erythrocytes were immunomagentically depleted from the blood. After this, CTCs were isolated using FACS based on fluorescence ($Venus^+$ cells). All of these cell populations were collected in lysis buffer for gene expression profiling. For all cell lines, primary tumor cells were collected. In addition, lung metastatic cells were collected from the 4T1, 66cl4 and 4T07 cell line and liver metastatic cells were collected from the 4T1 line. We were able to collect CTCs from the 4T1, 66cl4 and 4T07 cell lines (Table 3.4).

RNAseq libraries were generated from total RNA using the Nugen Ovation RNA-Seq System II. The libraries were sequenced on the Illumina platform (single-end, 50 cycles). Sequencing reads were mapped to the mm9 mouse genome using TopHat and differential expression was assessed using Cufflinks/Cuffdiff (Trapnell et al. 2010; 2009). The percentage of unique mappers to the genome was around 80% to 90% of total reads. For some lung samples, this was around 70%. Pearson correlation between biological replicates for each condition was above 0.85 for all samples, except for some lung samples (Fig. 3.5a).

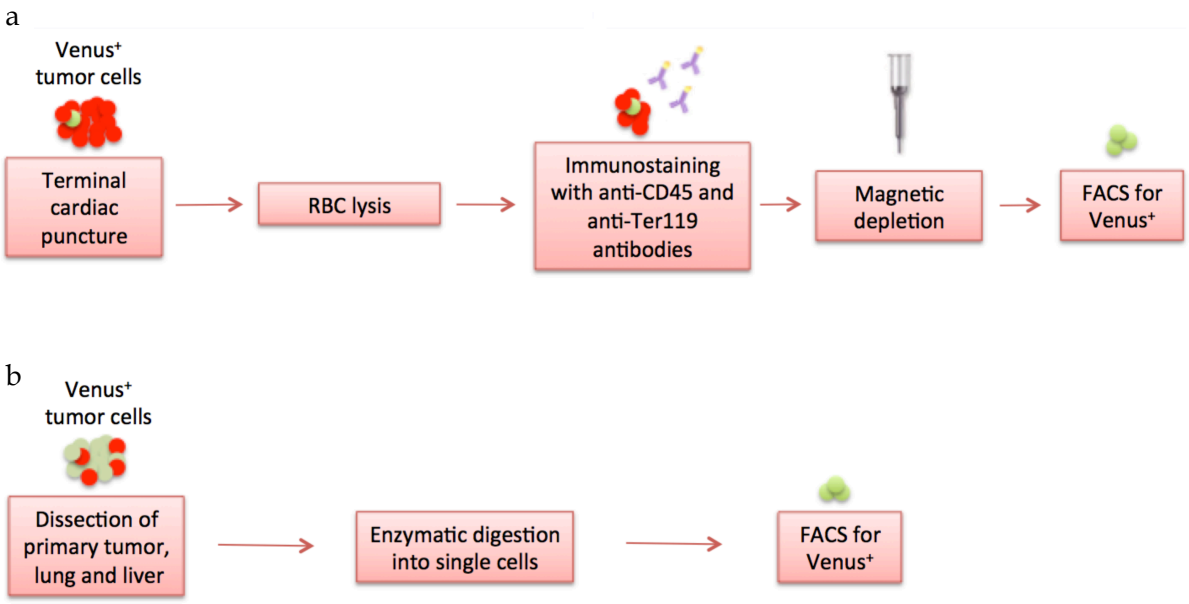


Figure 3.4 Isolation procedure for CTCs, primary and metastatic tumor cells
(a) Whole blood is isolated from mice through terminal cardiac puncture. After red blood cell (RBC) lysis, whole blood is stained with anti-CD45 (leukocytes) and anti-Ter119 (erythrocytes). The cells are immunomagnetically delpleted and Venus+ cells are isolated by FACS. (b) Primary tumor, lung and liver are harvested from mice and then enzymatically digested with collagenase and hyaluronidase into single cells. Venus+ cells are isolated by FACS.

Principle component analysis (PCA) on the RNAseq data revealed higher separation between each of the cell lines than separation between each of the replicates (not shown). Hierarchical clustering of the RNAseq data revealed a tropism based on the tissue origin (Fig 3.5b). All of the primary tumor cells clustered together and so did the CTCs and lung metastatic cells. Interestingly, the CTCs clustered closer to the primary tumors than to the lung metastatic cells.

In summary, we were able to obtain high-quality RNAseq libraries, even from samples with very low input RNA. The Pearson correlation between biological replicates were very high for most samples, except for some lung samples. This might be due to the harsh isolation procedure, in which we enzymatically digest the lung tissue into single cells. These lung samples also have a low percentage of unique mappers. Because of this, we decided to primarily focus on the analysis of the primary tumors and CTCs.

Another technical limitation is the sorting of these cells based on fluorescence. Because it is nearly impossible to get a pure population from the FACS sorts, there might be some contamination of non-tumorigenic cells. However, we can estimate that this percentage is less than 10%, since we were able to get over 90% pure populations from FACS sorts when we analyzed the post-sorted cell populations.

Because the tumor cells were tagged with a fluorophore, we were able to isolate the whole CTC population regardless of surface makers. The FDA-approved CellSearch system uses a mix of cytokeratins and EpCAM to identify and isolate CTCs (Müller et al. 2012). Since it is known that many tumor cells undergo EMT, in which they down-regulate epithelial makers, this system might not be able to isolate all CTCs in a given sample (Tsai and Yang 2013). A recent study showed that many CTCs from human cancer patients do show expression of mesenchymal markers (Yu et al. 2013). Fortunately, we are able to

Table 3.4 Number of FACS sorted cells collected for each of the 4T1 cell line

Cell line	Primary tumor [# of cells collected per mouse]	Lung mets [# of cells collected per mouse]	Liver mets [# of cells collected per mouse]	CTCs [# of cells collected per mouse]
67NR Venus	++ 200,000 cells	–	–	–
168FARN Venus	++ 200,000 cells	–	–	–
4T07 Venus	++ 200,000 cells	++ 10,000 – 20,000 cells	–	+ 1000 – 2500 cells
66cl4 Venus	++ 200,000 cells	++ 10,000 cells	–	+/– No cells – 30 cells
4T1 Venus	++ 200,000 cells	++ 20,000 cells	+ 50 – 250 cells	++ 8000 – 20,000 cells

n = 2 mice per group

a

	67NR	168FARN	4T07	66cl4	4T1
Tumor	0.99	0.88	0.99	0.97	0.98
CTCs	-	-	0.93	-	0.90
Lung	-	-	0.83	0.78	0.59
Liver	-	-	-	-	0.88

b

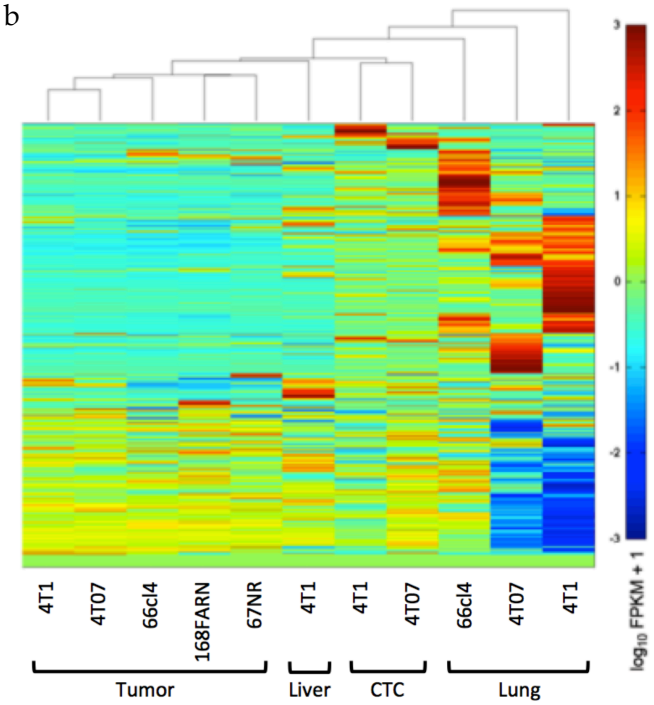


Figure 3.5 Statistical analysis and clustering of RNAseq data
(a) Pearson correlation between biological replicates for each of the cell line and the tissue that it was isolated from (b) Hierarchical clustering of RNAseq libraries. Columns represent averages of two mice.

circumvent this by using a fluorophore and only using a negative depletion, in which we removed leukocytes based on CD45 and erythrocytes based on Ter119.

Ptpn22 is a novel metastatic inducer

We wanted to identify potential metastatic regulators and thus looked for differentially expressed genes between the CTC population and the primary tumor. This was done in the case of the 4T1 and 4T07 cell lines. 12 genes were identified that were significantly overexpressed in the CTCs vs. primary tumor in both the 4T1 and 4T07 cell lines (above 4 fold and abundant > 10 FPKM, Fig. 3.6a,b). Some of these genes have been implicated with breast cancer and metastasis before. Aquaporin 1 (AQP1) is a water channel protein, whose expression has been found to correlate with aggressive basal-like breast carcinomas (Otterbach et al. 2010). Synuclein gamma (Sncg) is involved in perineural invasion and distant metastasis in a pancreatic cancer mouse model (Hibi et al. 2009). In addition, SNCG expression can be used as a marker for breast cancer prognosis, especially for stage III and IV breast cancer (Wu et al. 2007). G protein-coupled receptor kinase 4 (GRK4) has been identified to activate the MAPK signaling pathway in breast cancer cells (Matsubayashi et al. 2008). Because many of these candidate genes have been implicated with cancer progression and metastasis, we decided to screen these 12 genes in a pooled fashion.

In order to check the feasibility of an *in vivo* shRNA screen with these 12 genes, we decided to check the number of clones that engraft in the primary tumor and lung using a high complexity barcode library. This lentiviral barcode library was infected into 4T1 cells and were injected orthotopically into nude mice. After 4 weeks, primary tumor and lung were isolated and the overall number of barcodes were determined. On average, the

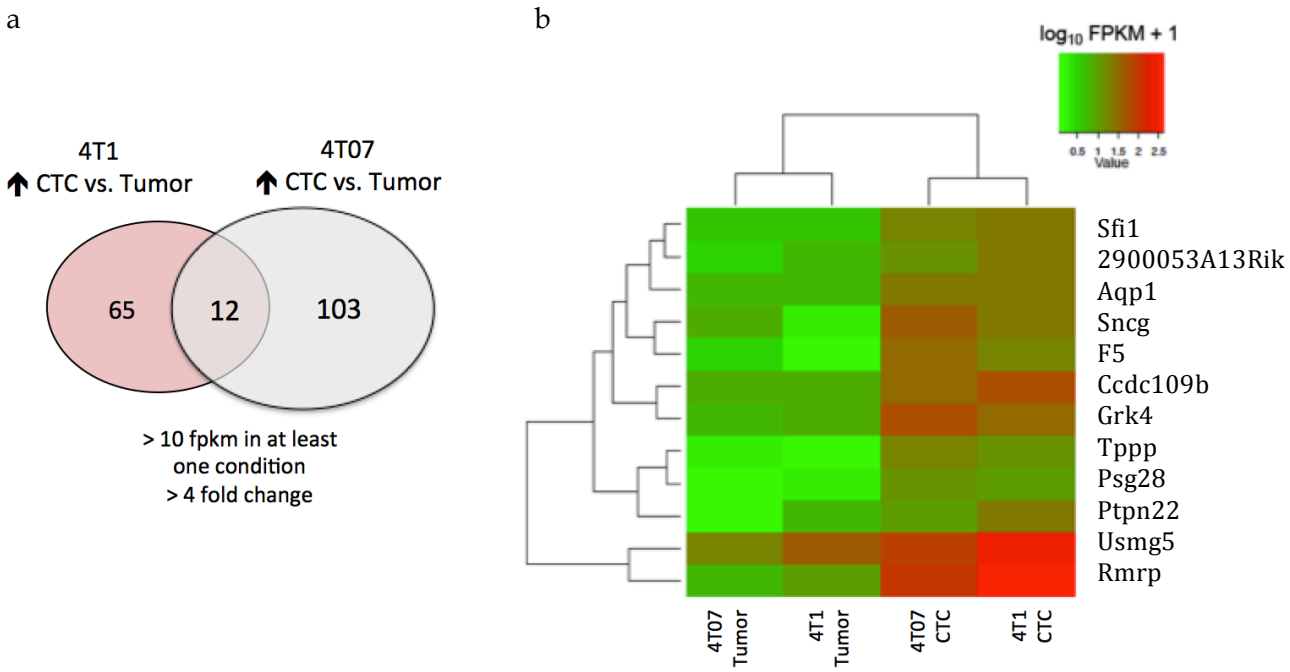


Figure 3.6 Overview of genes that are overexpressed in CTCs vs. primary tumor in 4T1 and 4T07

(a) Overlap of overexpressed genes between CTCs and primary tumor in both 4T1 and 4T07.
(b) Heatmap of the 12 overexpressed genes.

primary tumor contained around ~4700 barcodes and the lung contained around ~1550 barcodes. Because we wanted to have a ~100x fold representation of our shRNAs in the lung, we decided to use a maximum of 18 shRNAs per pool. Because of the unequal engraftment of some of the clones in the primary tumor and lung, we decided to use 6 mice per pool (see Chapter 4 for barcode representation of 4T1 cells).

We decided to screen these 12 candidate genes in an *in vivo* metastasis depletion assay (Fig 3.7a). For this, we infected 4T1 cells with shRNA pools (18 shRNAs per pool, 3-4 shRNAs per gene) and injected these cells orthotopically into the mammary gland of nude mice. After 4 weeks, we isolated the primary tumor and lung and subsequently sequenced shRNA representation. We scored shRNAs that were significantly depleted in the lung compared to the primary tumor using the software EdgeR (Robinson et al. 2010), since this would confirm that the candidate gene is necessary for metastasis. Ptpn22 was revealed as one of the top hits in this screen for metastatic inducers (Table 3.5).

Ptpn22 is a protein tyrosine phosphatase, which is expressed primarily in lymphoid tissue (Cohen et al. 1999). It has been implicated in autoimmunity susceptibility, response to injection and systemic inflammation (Brownlie et al. 2012). Interestingly, the 2012 TCGA Breast Cancer report highlighted Ptpn22 as one of the novel significantly mutated genes in their data set (Cancer Genome Atlas Network 2012).

Follow-up studies were performed with Ptpn22. Individual shRNAs against Ptpn22 and a control shRNA against Renilla were infected into 4T1 cells and each cell line was orthotopically injected into nude mice (Fig. 3.7b). Lung metastasis was evaluated after 4 weeks and we detected a reduction in lung metastases upon Ptpn22 knock-down (Fig 3.8a,b). Because of the promising results with Ptpn22, we decided to look at its potential mode of action during metastasis.

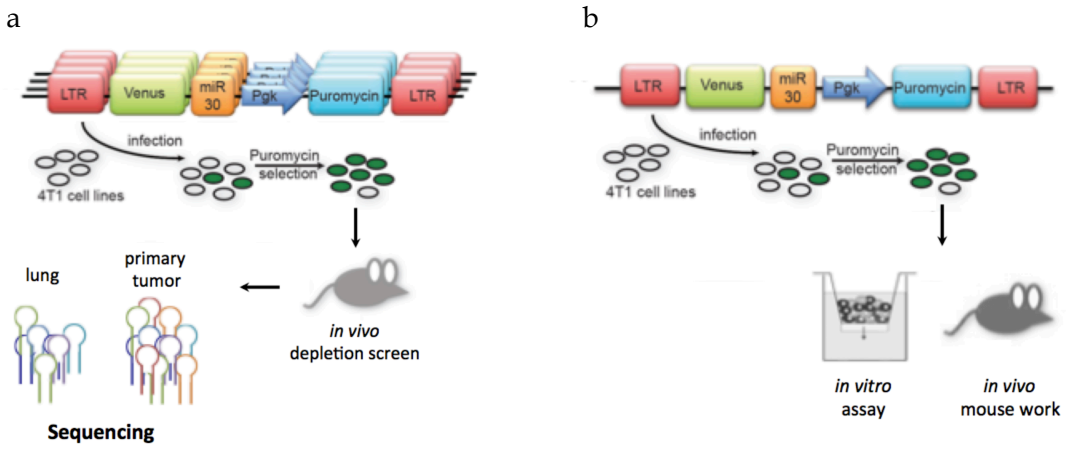


Figure 3.7 Schematic of pooled depletion metastasis screen and one-by-one validation
(a) Several shRNAs targeting multiple genes are pooled and infected into 4T1 cells. These cells are orthotopically injected into mice and the primary tumor and lung are harvested after 4 weeks. Depletion of specific shRNAs in the lung compared to the primary tumor confirms that this candidate gene is necessary for the metastatic process (b) A single shRNA targets a gene and is infected into 4T1 cells. These cells are used for *in vitro* experiments such as invasion assays and for *in vivo* assays, such as orthotopic injections.

Table 3.5: Depleted shRNAs from the *in vivo* metastasis screen

depleted shRNAs	log2 fold change	p value	FDR
Ptpn22_4	4.74	0.0011	0.0087
Ptpn22_1	2.59	0.0131	0.0474
Psg28_2	3.47	0.0037	0.0184
Psg28_3	3.04	0.0116	0.0474
Tppp_1	3.79	0.0043	0.0184
Sfi1_2	3.36	0.0056	0.0192
Ctrl (neutral)	1.14	0.2895	0.3990
Ctrl (neutral)	0.34	0.7463	0.9185

6 biological replicates per pool

In order to study the effects of Ptpn22 on metastasis formation, we infected 4T1 cells with Ptpn22 shRNAs and a control shRNA (Fig 3.7b) and carried out various *in vitro* and *in vivo* assays. We did not detect any changes in proliferation *in vitro*, as calculated by total cell number, and *in vivo*, as calculated by primary tumor growth, upon Ptpn22 knock-down in 4T1 cells (Fig 3.9a,b). In addition, there was no change in invasion in *in vitro* in a matrigel based invasion assay (Fig 3.9c). We used an shRNA against Sncg as a positive control (Jia et al. 1999). Because of these results, we were certain that there is likely no effect on proliferation and invasion upon Ptpn22 knockdown in 4T1 cells. We next drew our attention to CTCs and wanted to see if there is a change in numbers of CTCs. Interestingly, we were able to see a reduction in CTC numbers after orthotopic injection of 4T1 cells with Ptpn22 knock-down (Fig 3.9d). We used the same CTC isolation technique as for the gene expression profiling (Fig 3.4a). Because of the fact, that there was no change in invasion *in vitro*, but a change in the overall number of CTC counts in the blood, we hypothesized that CTC survival might be affected upon Ptpn22 knock-down.

We decided to use an *in vitro* system that mimics CTC survival, in which we can grow 4T1 cells in an anchorage-independent fashion on ultra-low attachments plates. In addition we used serum starvation to induce stress on these cells (Fig 3.10a). In these settings, we detected an increase of Ptpn22 expression in 4T1 cells (Fig 3.10b). Interestingly, there was a higher rate of apoptosis upon Ptpn22 knock-down in these cells compared to the control shRNA, as measured by an ELISA for mono- and oligonucleosomes in the cytoplasma (Fig 3.10c). This was also confirmed by looking at the gene expression signature (RNAseq) of these cells (not shown). Pathway analysis confirmed an enrichment of genes in "apoptosis" and "cell death" pathways and a reduction of genes in the "cell survival"

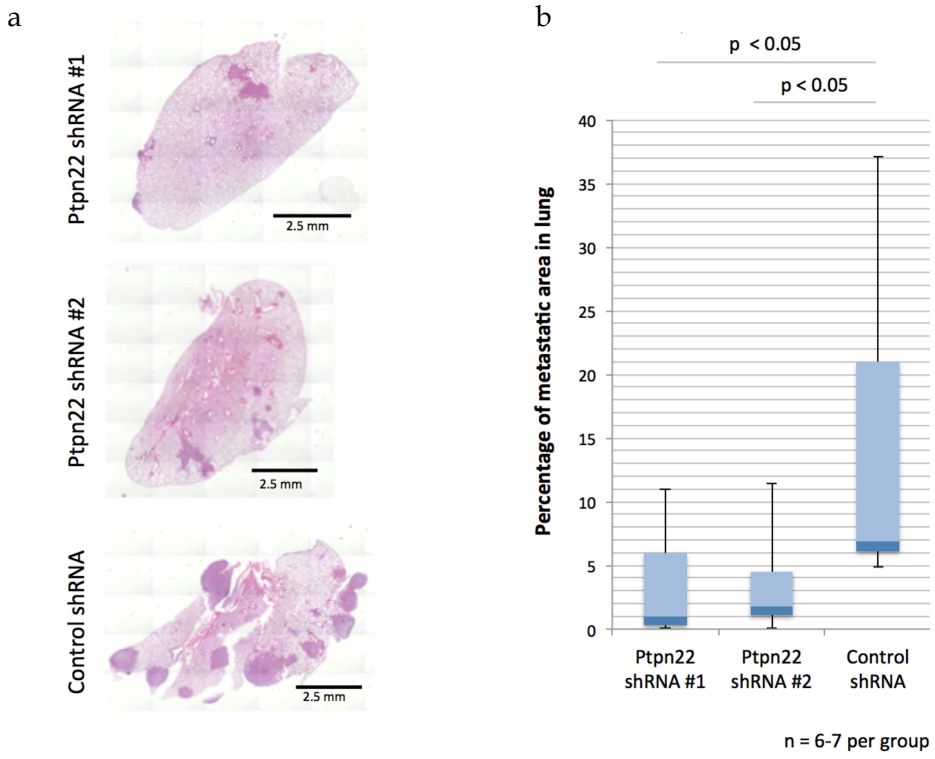


Figure 3.8 Ptpn22 knock-down reduced lung metastasis
(a) 4T1 cells infected with an shRNA against Ptpn22 or a control shRNA (Renilla) were each orthotopically injected into nude mice and lungs were harvested after 4 weeks. Metastatic lung burden was measured by H&E staining. (b) Quantification of the metastatic lung burden from (a). The graph represents box plots.

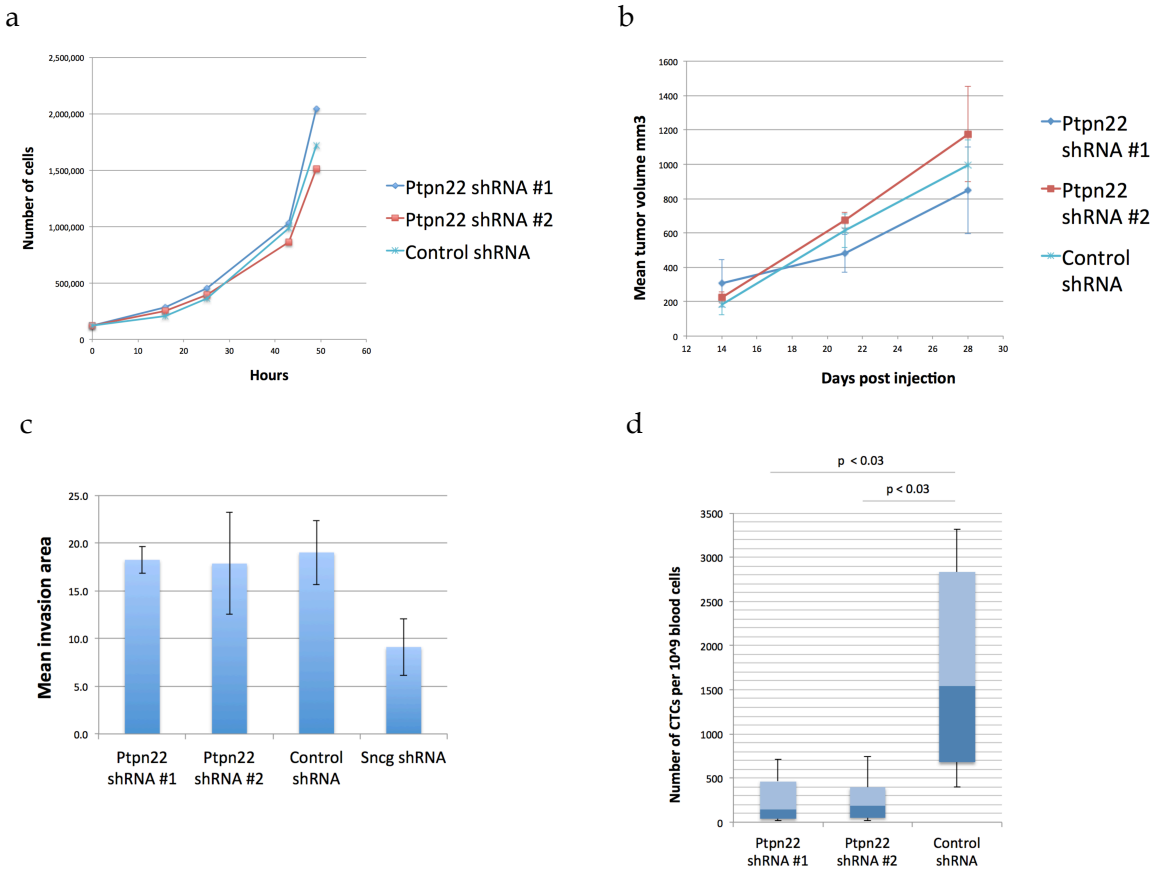


Figure 3.9 Ptpn22 knock-down reduced the number of CTCs, but does not affect proliferation and invasion

4T1 cells infected with an shRNA against Ptpn22 or a control shRNA (Renilla) were used for the following experiments. (a) No change in proliferation *in vitro* upon Pptn22 knock-down (b) No change in primary tumor growth upon Ptpn22 knock-down. (c) Matrigel-based invasion assays revealed no change in invasion upon Ptpn22 knock-down compared to the positive Sncg control. (d) 4T1 cells were orthotopically injected into nude mice and CTCs were harvested after 4 weeks. There was a reduction on the number of CTCs upon Ptpn22 knock-down. The graph represents box plots.

pathway. Thus, we think that overexpression of Ptpn22 in CTCs might protect these cells from stress-induced apoptosis (Fig 3.11).

In summary, we were able to show that Ptpn22 is overexpressed in the CTCs vs. the primary tumor when 4T1 or 4T07 cells are injected orthotopically into mice. The mode of action of Ptpn22 is likely on the survival of CTCs since there was no change in proliferation and invasion upon knock-down of Ptpn22 in 4T1 cells. Because we detected the change in CTC numbers, we wanted to establish an *in vitro* system that mimics the survival of CTCs in an anchorage-independent and stress inducing condition. Using this assay, we saw an increase in apoptosis upon knock-down of Ptpn22 in 4T1 cells, which then resulted in the decreased incidence of metastasis in the lung. Since Ptpn22 is a phosphatase, it will be interesting to detect specific substrates that might play a role in this mechanism. There are several phosphor-receptor tyrosine kinase and protein arrays that allow a semi high-throughput analysis of this.

The exact biological mechanism of how Ptpn22 might be pro-survival for the CTC population in the blood circulation is not known. Akt possesses pro-survival activities that can counteract apoptosis and thus it might be interesting to check the Akt phosphorylation levels with and without knock-down of Ptpn22 (Benbrook and Masamha 2011). In addition, it will be very important to validate the mode of action of Ptpn22 on lung metastasis in human cell lines such as the claudin-low/basal MDA-MB-231 and MDA-MB-436 cell lines. Likewise, it will be interesting to see if we could increase the incidence of lung metastasis by over-expressing Ptpn22 such as in the 4T07 cell line or in the previously mentioned human cell lines.

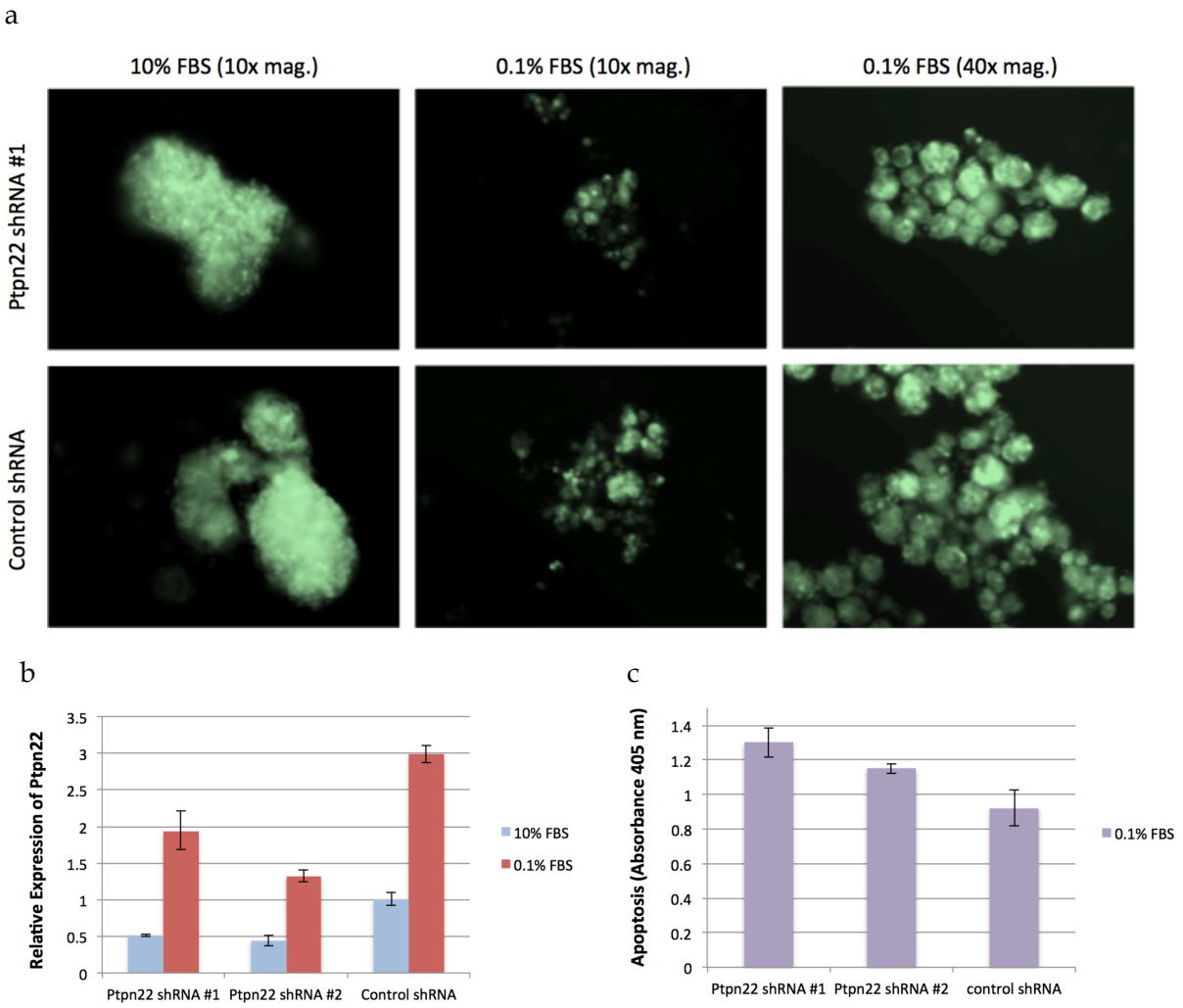


Figure 3.10 Ptpn22 knock-down increased rate of apoptosis *in vitro*
(a) 4T1 cells grow in organoids on ultra-low attachment plates under normal conditions (10% FBS) and under serum starvation (0.1% FBS) mimicking CTC survival (b) 4T1 cells display an up-regulation of Ptpn22 when these cells are grown on ultra-low attachment plates and serum starved (c) Under these conditions, there is an increase in apoptosis upon Ptpn22 knock-down (measured by ELISA for mono- and oligonucleosomes in the cytoplasm)

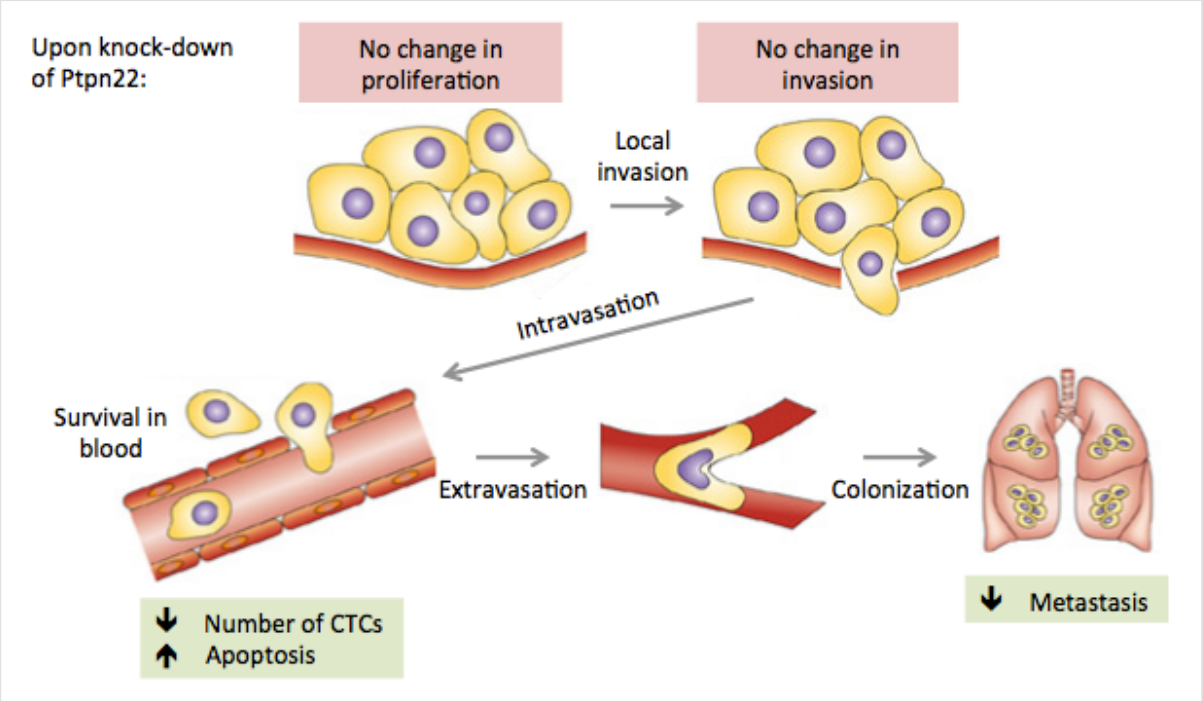


Figure 3.11 Mechanistic effect of Ptpn22 on metastasis
Knock-down of Ptpn22 in 4T1 cells does not affect proliferation and invasion, but results in a reduction on the number of CTCs in the blood stream and a higher apoptotic rate when these cells are subjected to stress. This results in an overall reduction of metastasis to the lung.

Methods

Cell culture

The mouse mammary tumor cell line 4T1 was purchased from ATCC and the 66cl4, 4T07, 168FARN and 67NR cell lines were a kind gift from Fred Miller. These cell lines were cultured in DMEM high glucose (Life Technologies) supplemented with 5% fetal bovine serum (Thermo Scientific), 5% fetal calf serum (Thermo Scientific), non-essential amino acids (Life Technologies) and penicillin streptomycin (Life Technologies).

Virus production

All retroviral vectors were packaged using Platinum-A packaging cells. The lentiviral barcode library was packaged using 293-FT lentivirus packaging cells.

Animal studies

All mouse experiments were approved by the Cold Spring Harbor Animal Care and Use Committee. All orthotopic injections were performed using 2 x 10^5 mouse mammary tumor cells resuspended in 20 ul of a 1:1 mix of PBS and growth factor reduced Matrigel (BD Biosciences). Cells were injected into mammary gland #4 of 8 week-old female nude or BALB/c mice (JAX). Tumor, lung and blood were harvested from mice for further processing. Primary tumor volume was measured using the formula $V = 1/2(LxW^2)$, where L is length and W is width of the primary tumor.

$CD4^+$ and $CD8^+$ depletion

Anti-CD4 as well as anti-CD8 antibodies were produced by the CSHL Antibody Core facility using the following hybridome lines: GK1.5 for $CD4^+$ T cell depletion, which is

an IgG2b-producing rat T cell hybridoma line (Dialynas et al. 1983) and 53-6.72 for $CD8^+$ T cell depletion, which is an IgG2a-producing rat hybridoma that reacts with both the Lyt2.1 and Lyt2.2 alleles (Ledbetter and Herzenberg 1979).

BALB/c animals were intraperitoneally injected with 200 µg of each antibody every 5 days. PBS was used as a negative control. The spleen of a BALB/c mouse treated with anti-CD4 and anti-CD8 antibodies was isolated and mashed on a 60 µm cell strainer. Single cells were centrifuged and resuspended in PBS before staining with CD4-Alexa648 (BD Biosciences) and CD8-FITC (Miltenyi Biotec) antibodies.

Quantification of lung metastatic burden

The lung metastatic burden was evaluated in five-micron sections stained with a standard H&E protocol. Quantification was performed using ImageJ Software (NIH). The total lung area and metastatic area was determined by tracing the individual regions by hand. The relative metastatic area was then calculated.

Isolation of blood, primary tumor and lung and liver metastatic cells

Tumor, lung and liver were enzymatically digested into single cells using a collagenase and hyaluronidase buffer (Stemcell) and then sorted based on fluorescence using the FACS AriaII (BD Biosciences). Whole blood was collected by terminal cardiac puncture, erythrocytes were removed by red blood cell lysis (ACK Buffer, Sigma) and then $CD45^+$ leukocytes and $Ter119^+$ erythrocytes (BD Biosciences) were immunomagentically depleted from the blood using LS columns (Miltenyi Biotec). After this, CTCs were isolated using FACS based on fluorescence using the FACS AriaII (BD Biosciences).

Complexity analysis

Genomic DNA was isolated using phenol chloroform extraction. The barcodes of the lentiviral library were amplified using a one-step PCR protocol. For each sample, 96 individual PCR reactions of 500 ng of genomic DNA were carried out using Phusion (Thermo Scientific). The PCR was carried out for 33 cycles. PCR products were purified using the PCR purification kit (Qiagen). PCR products were size selected on an agarose gel, and then sequenced on the Illumina GaIIx sequencer generating 36 nt single-end (SE) reads. The complexity library was analyzed in collaboration with Dr. Simon Knott.

shRNA analysis

Genomic DNA was isolated using phenol chloroform extraction. The shRNAs were amplified using two-step PCR protocol. The first PCR was carried out for 25 cycles. PCR products were purified using the PCR purification kit (Qiagen). The second PCR was performed using 500 ng of PCR product from the first PCR. The second PCR was carried out for 25 cycles and PCR products were again purified using the PCR purification kit (Qiagen). PCR products were size selected on an agarose gel and sequenced on the Illumina GaIIx sequencer generating 36 nt single-end (SE) reads. shRNA depletion (fold change between tumor to lung) was calculated using EdgeR with a FDR threshold of <0.05.

RNA-Seq library preparation

Total RNA was purified and DNAse treated using the Qiagen RNeasy Mini Kit. RNA integrity (RNA Integrity score > 9) and quantity was measured on an Agilent Bioanalyzer (RNA Nano kit). The NuGEN Ovation RNA-Seq V2 protocol was carried out on 100 ng of total RNA. cDNA was fragmented using a Covaris sonicator according to the

manufacturer's instruction to yield a target fragment size of 300 bp. The fragmented cDNA was subsequently processed through end-repair, A-tailing and adapter ligation using our in-house one-shot low-input protocol. A minimal PCR amplification was followed using Phusion (Thermo Scientific) for 16 cycles. Each sample was sequenced on the Illumina GaIIx sequencer generating 50 nt single-end (SE) reads. Sequencing reads were mapped to the mm9 mouse genome using TopHat and differential expression was assessed using Cufflinks/Cuffdiff.

Exome capture library preparation

Genomic DNA was purified using phenol-chloroform extraction. Then, the genomic DNA was fragmented using a Covaris sonicator according to the manufacturer's instruction to yield a target fragment size of 300 bp. The fragmented cDNA was subsequently processed through end-repair, A-tailing and adapter ligation using our in-house one-shot low-input protocol. A minimal PCR amplification was followed using Phusion (Thermo Scientific) for 16 cycles. Then, the SeqCap EZ Mouse exome capture (Roche NimbleGen) was used according to the manufacturer's instruction and then sequenced on the Illumina HiSeq sequencer generating 76 nt single-end (SE) reads. The Exome-capture sequencing was analyzed in collaboration with Prof. Molly Hammell (CSHL).

qRT-PCR

Total RNA was purified and DNAse treated using the RNeasy Mini Kit (Qiagen). Synthesis of cDNA was performed using SuperScript III Reverse Transcriptase (Sigma). Quantitative PCR analysis was performed on the Eppendorf Mastercycler ep realplex. All signals were quantified using the ΔCt method and were normalized to the levels of Gapdh.

Proliferation rates

Proliferation assays were performed by counting viable cells over 72 hours.

Invasion Assay

Invasion assays were done using BioCoat Matrigel invasion chambers (BD Biosciences). 2 x 10^5 4T1 cells were plated per well and FBS was used as a chemoattractant. After 18 hours, the invaded cells were stained with 5 µM SYTO 60 (Life Technologies). Chambers were scanned using the Odyssey system (Licor).

Apoptosis Assay

The Cell Death Detection ELISA (Roche) was carried out in duplicates according to the manufacturer's instructions.

Chapter contributions

Rebecca Berrens assisted with RNAseq library preparation and analysis.

Chapter 4: A model of tumor heterogeneity reveals vascular mimicry as a driver of metastasis

Metastasis requires that primary tumor cells evolve the capacity to intravasate into the lymphatic or cardiovascular systems, extravasate into a target organ, and colonize secondary sites that are physiologically distinct from that of the primary (Nguyen et al. 2009; Vanharanta and Massagué 2013). We have developed a mouse model to probe the role of breast tumor heterogeneity in multiple stages of disease, from primary tumor growth to the establishment of metastases. We found that distinct clones display specialization, for example dominating the primary tumor, contributing to metastatic populations, or showing tropism for entering the lymphatic or vasculature systems. We correlated these stable properties to distinct gene expression profiles. Those clones that efficiently entered the vasculature expressed two secreted proteins, Serpine2 and Slpi, that were necessary and sufficient to program these cells for vascular mimicry. We hypothesize that vascular mimicry drives the ability of some breast tumor cells to contribute to distant metastases while simultaneously satisfying a critical need of the primary tumor to be fed by the vasculature. Enforced expression of SERPINE2 and SLPI in human breast cancer cell lines also enabled vascular mimicry, and SERPINE2 and SLPI were overexpressed preferentially in human patients that had metastatic relapse. Thus, these two secreted proteins, and the phenotype they promote, may be broadly relevant as drivers of metastatic progression in human cancer.

4T1 as a tumor heterogeneity model

Tumor heterogeneity has been posited as a significant contributor to the failure of cancer therapy (Meacham and Morrison 2013; Ding et al. 2012; Mullighan et al. 2008; Lohr et al. 2014). Two models have been proposed to explain the heterogeneity in genotype and phenotype that is observed within human cancers. The so-called cancer stem cell model purports that a fraction of cells within a cancer have high tumor forming potential (Beck and Blanpain 2013). These fuel the production of bulk tumor cells that are more differentiated and thus less likely to engraft in transplantation experiments or to contribute to therapy resistance or metastasis. The clonal evolution model invokes progressive Darwinian selection of genetically distinct clones with diverse properties, for example resistance to therapy, predisposition for dissemination to secondary sites, and the ability to produce metastases (Greaves and Maley 2012). While it is nearly certain that human tumors contain populations of cells that are phenotypically and genetically distinct, neither of the two most widely considered models likely captures fully the mechanisms that lead to heterogeneity or its consequences for human disease (Shackleton et al. 2009). For example, discrete cell populations in human tumors can act symbiotically, with one secreting growth factors on which the other depends (Neiva et al. 2014; Grivennikov et al. 2009). Moreover, tumors and cancer-associated stromal populations can act together to promote tumor progression and therapy resistance (Orimo et al. 2005; Olive et al. 2009). Thus both the mechanisms that generate tumor heterogeneity and its impacts on cancer therapy are likely more complex and subtle than is generally envisaged.

To date, the most detailed studies of tumor heterogeneity have been retrospective (Mullighan et al. 2008; Lohr et al. 2014; Ding et al. 2010; Navin et al. 2011). For example, single cell analyses of human breast tumors have illustrated evolutionary paths of genetic

diversification (Navin et al. 2011). In such cases, genetic variation could not be associated with differences in the behavior and capabilities of clonal populations and their specific contributions to disease. We, therefore, wished to complement such studies by creating an experimental model of tumor heterogeneity. Our goal was to enable correlation between variations in genotype and gene expression with cellular behaviors that contribute to tumor development and metastatic progression.

Identifying the factors required to overcome each step of the metastatic process has been difficult, as comparative genomic studies have been performed mainly with cells from the primary and secondary lesions, preventing stratification of the requirements for intravasation vs. extravasation (Bos et al. 2009; Minn et al. 2005). One of the most widely applied models used to study metastasis is transplantation of the 4T1 cell line into the mammary fat pad of mice (Dexter et al. 1978). The 4T1 cell line can be engrafted into syngeneic or immunocompromised recipients. Orthotopic transplantation into the mammary fat pad leads both to the development of primary tumors and to outgrowth of metastases in lymph nodes and secondary organ sites that are accessible via the vascular system. Here, we wished to apply this model to connect heterogeneity with metastasis, through the tracking, via next generation sequencing (NGS), of individual molecularly barcoded 4T1 clones throughout the various stages of metastatic disease progression. Such a system allows individual cells to be assessed for their ability to intravasate and extravasate, which, in turn, should allow the factors driving each step to be ascertained. In addition, we sought to determine whether each cell within the 4T1 population had equivalent properties or whether certain subsets of cells might display stable differences in behavior that could be studied in clonal populations.

Clonal analysis of 4T1 transplantation

To address this question, we marked individual cells with a molecular barcode via lentiviral infection (Fig. 4.1a and 4.2a). We drew from a complex mixture, five different cohorts of 100,000 cells each and introduced these into immunocompromised recipients. These were chosen because the barcode library encodes mCherry and hygromycin resistance, and we wished to avoid selective pressures imposed by immune rejection based upon foreign antigens. After 24 days, primary tumors, brachial lymph nodes, blood, lungs, livers, and brains were collected, and the barcode populations within each tissue were quantified via next generation sequencing (NGS). Though each engrafted population was distinct, there was a set of clones (~1,400) that was present in all samples. We asked whether these clones showed consistent behavior across all 5 experiments.

An analysis of the distributions of clone abundances across tissues indicated that only a subset of engrafted cells were able to intravasate into the cardiovascular and lymphatic systems, and an even smaller set were able to extravasate into the lung, brain and liver (Fig. 4.3a). A two-dimensional clustering of the clone distributions across organs resulted in all tumor, blood and lymph-node samples forming separate groups (Fig. 4.1b). Although the blood-born metastases samples formed one large meta-cluster, individual samples failed to fully group together based on their organ of origin.

Two conclusions were drawn from this analysis. First, clone abundance within the primary tumor did not correlate with abundance in CTCs or secondary lesions (Fig. 4.3b). Second, distinct groups of clones contributed to lymph node and blood-born metastases. Highly significant overlap existed between abundant clones in the blood-born metastases, and between these and CTCs (p-value < 0.001). However, no significant overlap was observed when comparing these sets to the abundant clones in the lymph node (Fig. 4.3c).

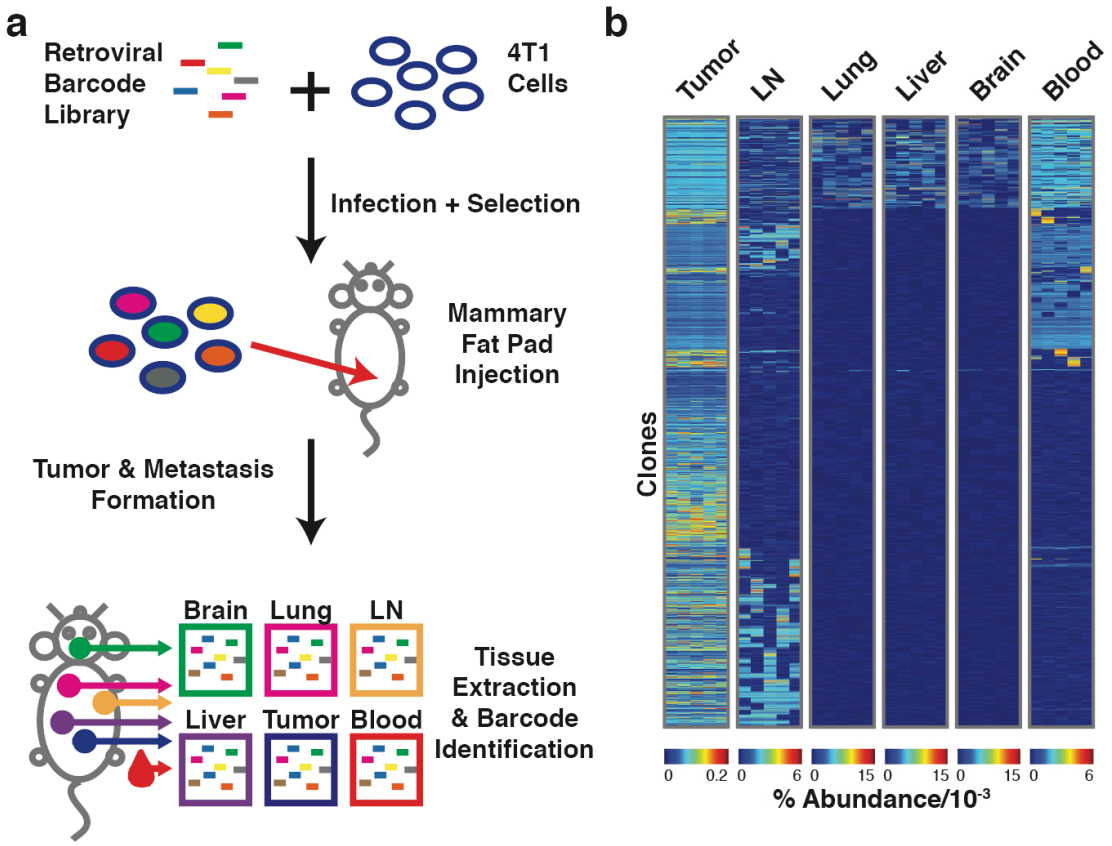


Figure 4.1. Clonal analysis of 4T1 transplantation by molecular barcoding
(a) Retroviral barcoding strategy for identifying clonal populations within the 4T1 cell line. (b) Relative proportions of clones that engrafted in all animals in the lymph-node (LN), lung, liver, brain and blood. Each column represents an independent experiment. Shown are the ~1400 clones that successfully engrafted in all animals.

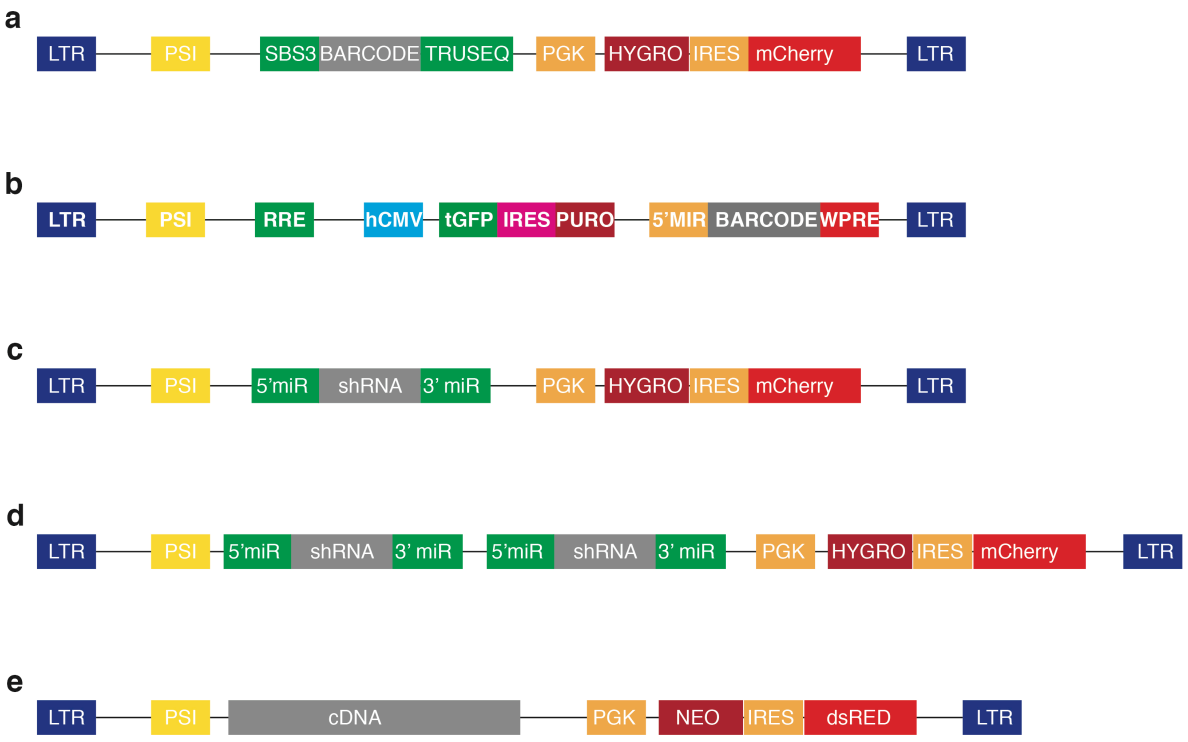


Figure 4.2: Overview of vector plasmids
(a) Schematic of the retroviral barcode vector. b. Schematic of the lentiviral barcode vector. (c) Schematic of the retroviral shRNA vector used for single gene knockdown. (d) Schematic of the tandem retroviral shRNA vector used for double gene knockdown. (e) Schematic of the retroviral cDNA vector used for gene overexpression.

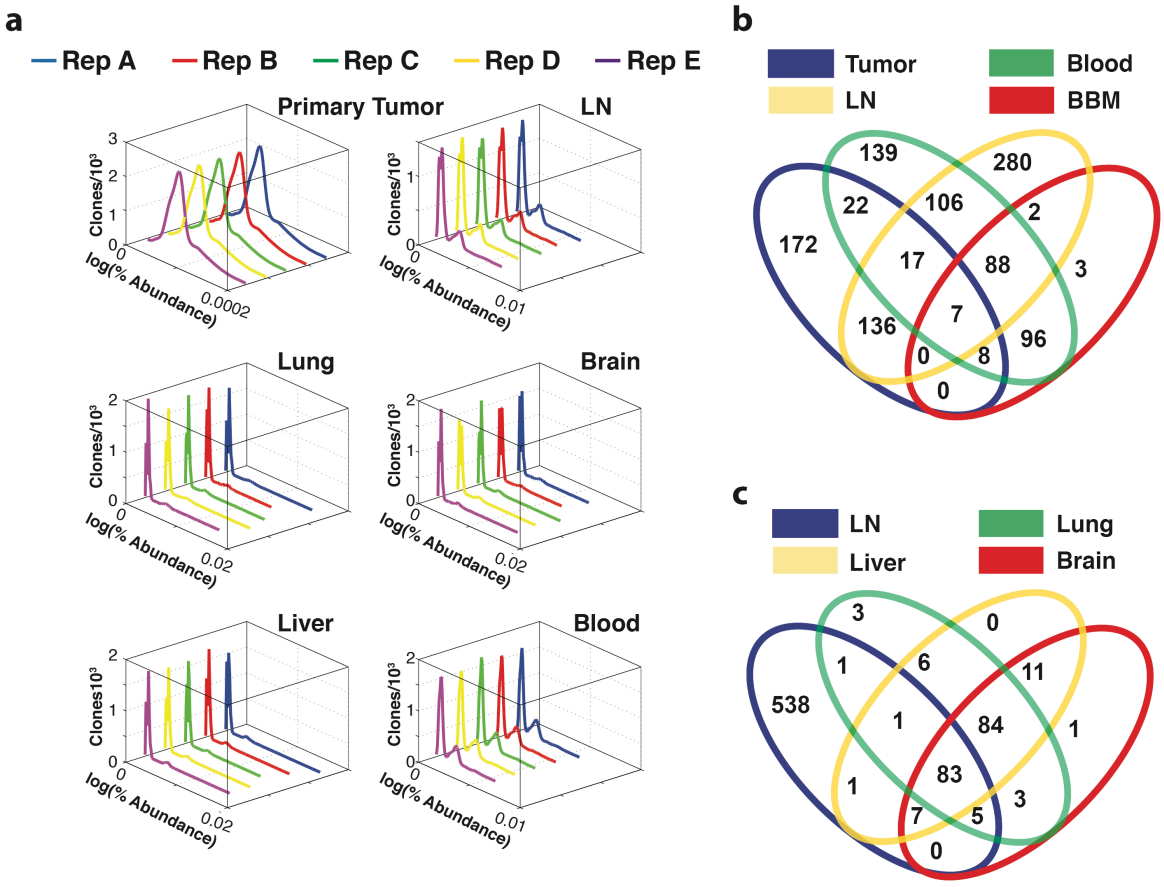


Figure 4.3: Clonal abundances in the primary tumor and secondary sites after 4T1 transplantation

(a) Distribution of clone abundances at the primary tumor and secondary sites for the 4T1 clones that engrafted and contributed to tumor formation in all five animals upon orthotopic injection. (b) Overlap of abundant clones in the primary tumor, lymph-node (LN), blood and in all blood born metastases (lung, liver and brain). (c) Overlap of the abundant clones in the LN, liver, lung and brain.

These studies implied that clonal populations within the 4T1 line have acquired stable differences that enable them to contribute to different aspects of disease progression.

Focused analysis of a subset of 4T1 clones

In order to gain further insight into these phenomena, using single cell FACS sorting, we established clonal lines from 4T1 cells that had been infected with a lentiviral barcode library (Fig. 4.2b and 4.5a). Twenty-three lines were chosen that stratified the various growth rates and morphologies observed during cell line establishment (hereon referred to as 4T1-A through W, Fig. 4.4). After minimal propagation, these lines were pooled in equal amounts and orthotopically injected into mice. In addition, the pool was propagated *in vitro*. After 14 days, primary tumors were removed from one cohort of mice and aliquots of the *in vitro* samples collected. After 24-days, a second set of *in vitro* aliquots were acquired and the primary tumors, brachial lymph nodes, blood, lungs, livers and brains were extracted from the remaining animals. The abundance of each clone in each sample was determined by NGS.

Figure 4.5b depicts the clonal composition of each primary tumor and *in vitro* pool after 14 and 24 days. A comparison of samples extracted at 14 days shows high concordance between the *in vitro* samples and the primary tumor clone compositions. However, while the *in vitro* composition is maintained after 24 days, the tumor compositions are altered such that all tumors are composed, primarily, of clone 4T1-I. 4T1-I's accelerated proliferation was subsequently recapitulated when the lines were, individually, orthotopically injected and it showed the greatest corresponding differential of tumor volumes between 14 and 24 days (Fig. 4.6a).

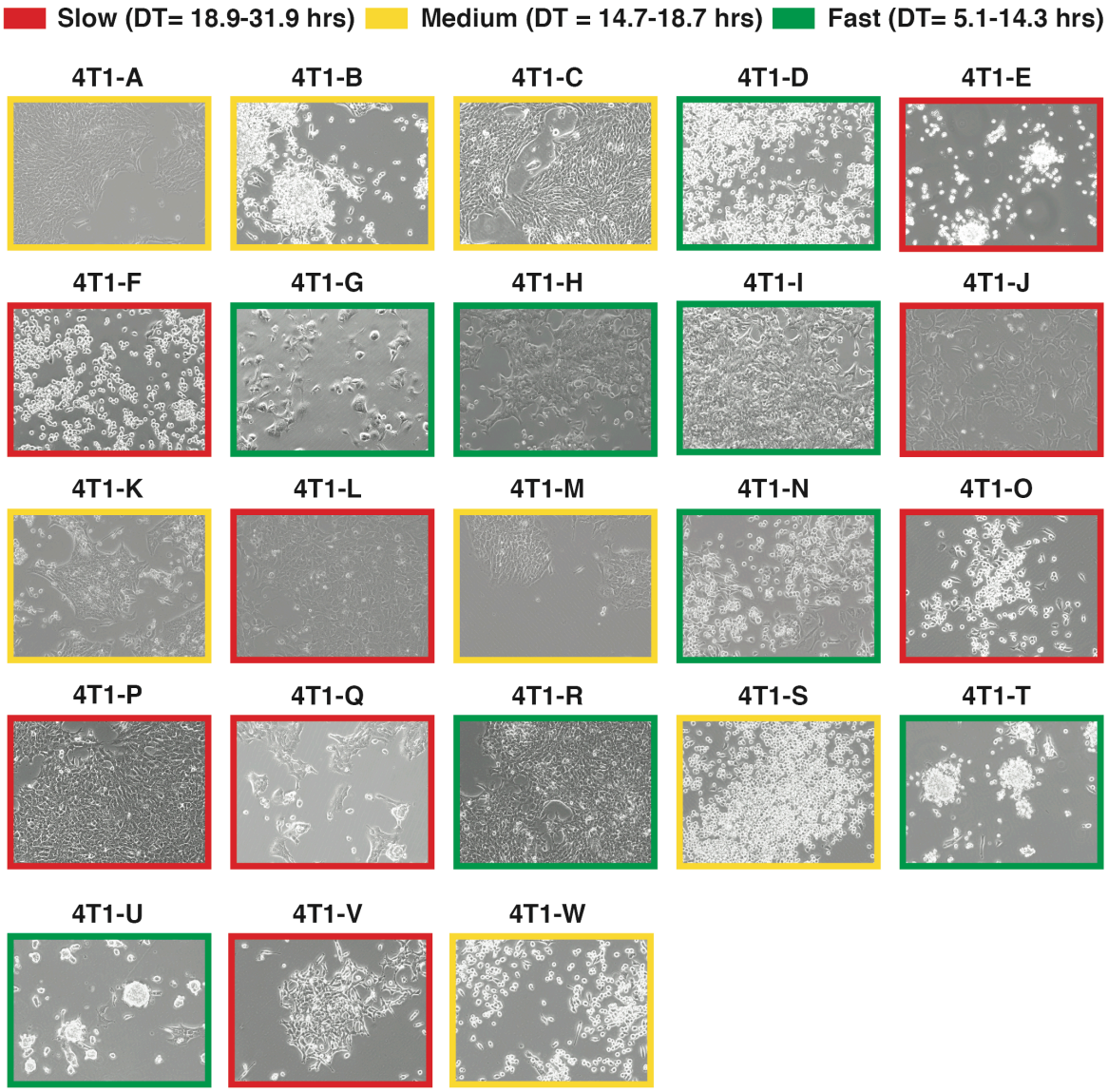


Figure 4.4: Morphology and proliferation rates of all clonal lines
10X microscopic phase-contrast images of the 23 clonal lines. Doubling times (DT), as calculated over a three-day period, are color coded in the image borders.

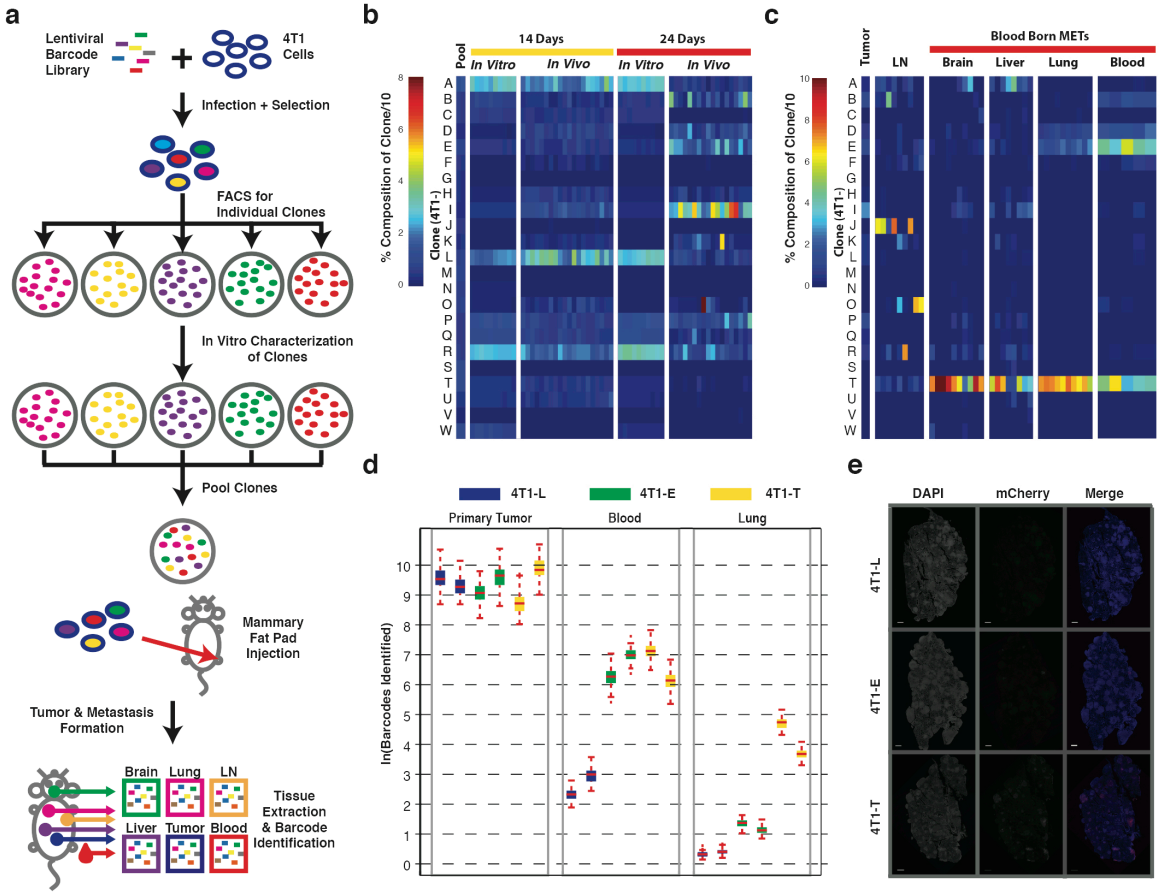


Figure 4.5. Focused analysis of a subset of 4T1 clones throughout metastatic disease progression

(a) Strategy for isolating individual molecularly barcoded 4T1 clones. (b) A comparison of the relative proportions of the isolated clonal lines *in vitro* and in orthotopic primary tumors at 14 and 24 days. Columns represent biological replicates. (c) A comparison of clonal line proportions amongst the CTCs in the blood as well as the secondary lesions in lymph-node (LN), brain, liver and lung of animals corresponding to the tumors extracted at 24 days described in (b). (d) Subclonal analysis of 4T1-L, -E and -T cells via secondary barcode library infection. Each clonal line was separately infected with a second barcode library, pooled with the 22 other lines and then analyzed for sub-clonal populations via analysis of the secondary library. (e) Immunoflourescence analysis for mCherry in lung metastases resulting from each of the three pooled injections discussed in (d). mCherry is expressed in the secondary barcode library.

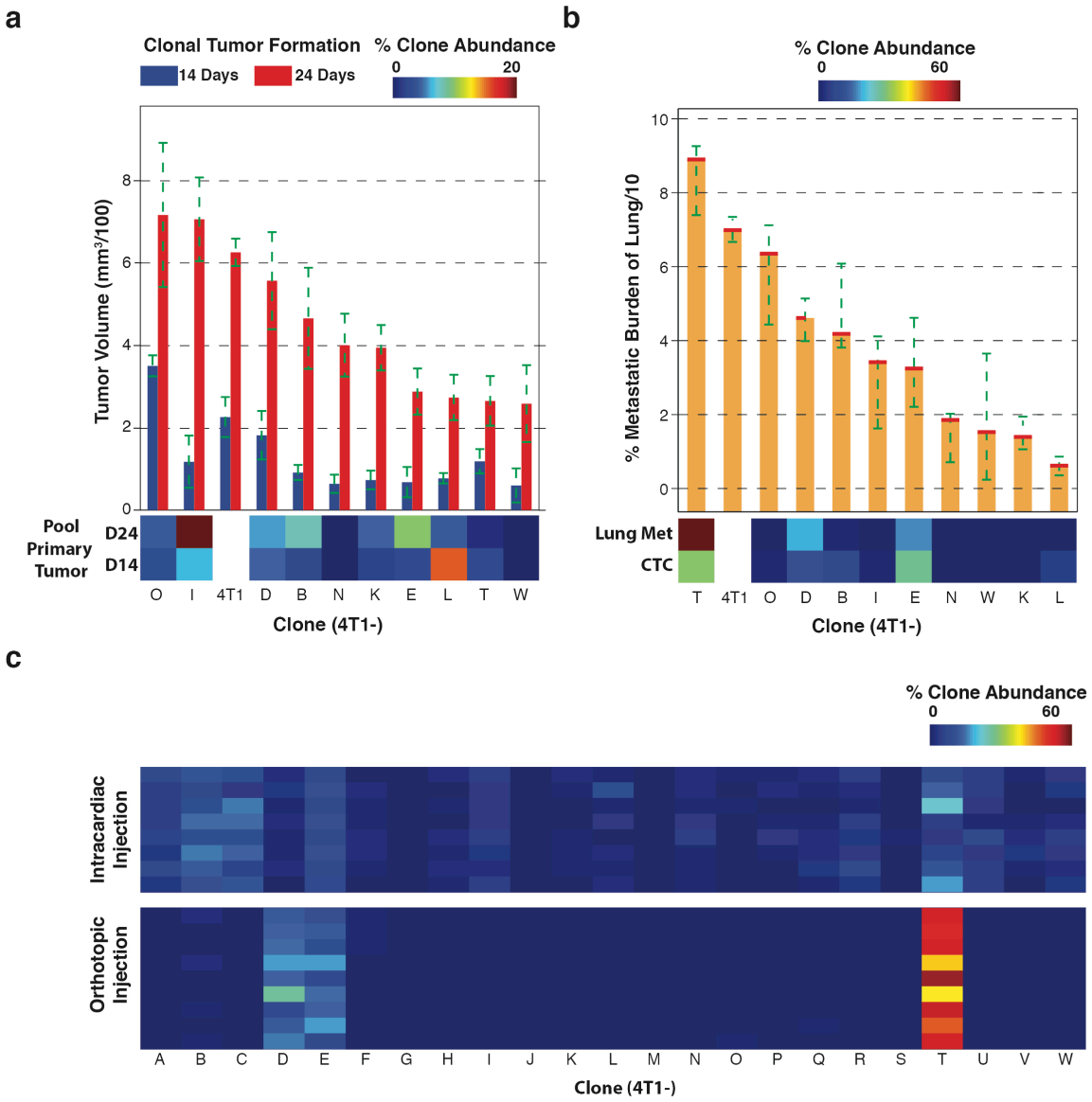


Figure 4.6: Tumor growth and metastases rates for individually injected clones
(a) Primary tumor volumes resulting from orthotopically injected individual clonal cell lines. Measurements were taken 14 and 24 days after injection. (b) The percentage of lung metastatic burden for the animals discussed in (a) 24 days after injection. (c) A comparison of the clonal composition of lung metastases when the pool of 23 clonal lines was orthotopically injected (bottom panel) vs. injected into the left cardicac ventricle of mice (top panel).

Figure 4.5c depicts the composition of the blood and secondary lesions, and one can see that this small system recapitulates the results of the large-scale trial discussed in Figure 4.1. First, a clone's abundance in the primary tumor does not correlate with its capacity to intravasate the cardiovascular system or form secondary lesions (e.g. 4T1-I is absent from all secondary sites). Second, the ability of individual clones within the population to intravasate the cardiovascular system is a driving force in determining the final population of blood-born metastases and in producing clonal tropism between the lymph-node and blood-born metastases. This is striking in this system where 4T1-E and -T dominate the CTC population, only 4T1-T is present at blood-born secondary lesions, but neither is present in the lymph-node. 4T1-T's proclivity to form blood-born metastases was further confirmed when the lung metastatic burden of the mice discussed for Figure 4.6a was quantified (Fig. 4.6b). In addition, the effect that an intravasation step has on the heterogeneity of secondary lesions was further confirmed when the pool of clones was injected into the left cardiac ventricle of mice. Under these conditions an entirely different clonal distribution is observed at the secondary lesion (Fig. 4.6c).

To assess how the complexity of representative lines are affected at each selection stage, we infected each of the clonal lines 4T1-E, -L and -T with a second barcode library (Fig. 4.1a), pooled them separately with the remaining 22 lines and, separately, orthotopically injected the resultant three pools. The number of engrafted 4T1-E, -L and -T sub-clones was found to be similar (Fig. 4.5d). However, selection for the ability to intravasate the cardiovascular system resulted in a significant reduction in the number of 4T1-L clones in the blood (p-value < 0.001). Finally, selection for the ability to extravasate into the lung resulted in 4T1-E being significantly reduced in clone-number (p-value < 0.001). Only 4T1-T subclones gave rise to abundant lung metastases. Though we observe some loss in

complexity at each stage of the metastatic process, the proclivities of each line are general properties of many, if not all, cells within the clone. This suggested that profiling these populations could provide clues to the mechanistic basis of their differential abilities. The results of the combined selection steps are depicted visually in Figure 4.6e, where lungs corresponding to the three pools were stained for mCherry, which is expressed in the second barcode library.

The chromosomal integration site of some of the barcodes including 4T1-E, -L and –T were evaluated and are listed in Table 4.1. Most of the barcodes were inserted outside of genes. The 4T1-T barcode was inserted in the Outer dense fiber protein 2-like isoform 1/2. There are no indications in any publications that this gene is important for tumor or metastasis development.

Transcriptional profiling of clonal cell lines

A qualitative analysis of the clonal lines revealed that, while the majority of cells displayed clear epithelial or mesenchymal morphologies, 4T1-E and 4T1-T (along with 4T1-U) displayed an "organoid" morphology (Fig. 4.4, p-value = 0.0119), hinting that clones 4T1-E and -T share characteristics that promote intravasation. To ascertain these drivers, we performed RNA-seq on all lines and identified genes significantly overexpressed in 4T1-E and -T as compared to at least half of the remaining lines. In addition, RNA-seq was performed on matched primary tumors and lung metastases to identify genes significantly up-regulated at secondary lesions. These gene sets were then intersected to define a set of 12 genes associated with intravasation into the cardiovascular system (Fig. 4.7a).

Table 4.1: Chromsomal integration sites of 4T1 clonal cell lines

Clone (4T1-)	Chromosomal Integration site
B	Within Early endosome antigen 1
E	68586 bp at 5' side: interferon epsilon precursor 188052 bp at 3' side: S-methyl-5'-thioadenosine phosphorylase
I	Within RNA-binding protein 40
K	78837 bp at 5' side: Rho GTPase-activating protein 21 isoform 1132146 bp at 3' side: Enkurin
L	16181 bp at 5' side: Cell division control protein 42 homolog isoform 2 43229 bp at 3' side: Uncharacterized protein LOC242711 precurdor
O	15527 bp at 5' side: Uncharacterized protein LOC101056086 2813 bp at 3' side: Cadherin-11-like
R	95328 bp at 5' side: Type-2 angiotensin II receptor 123447 bp at 3' side: Amino acid transporter ATB0+
T	Within Outer dense fiber protein 2-like isoform 1/2
W	Within Myb-related protein B

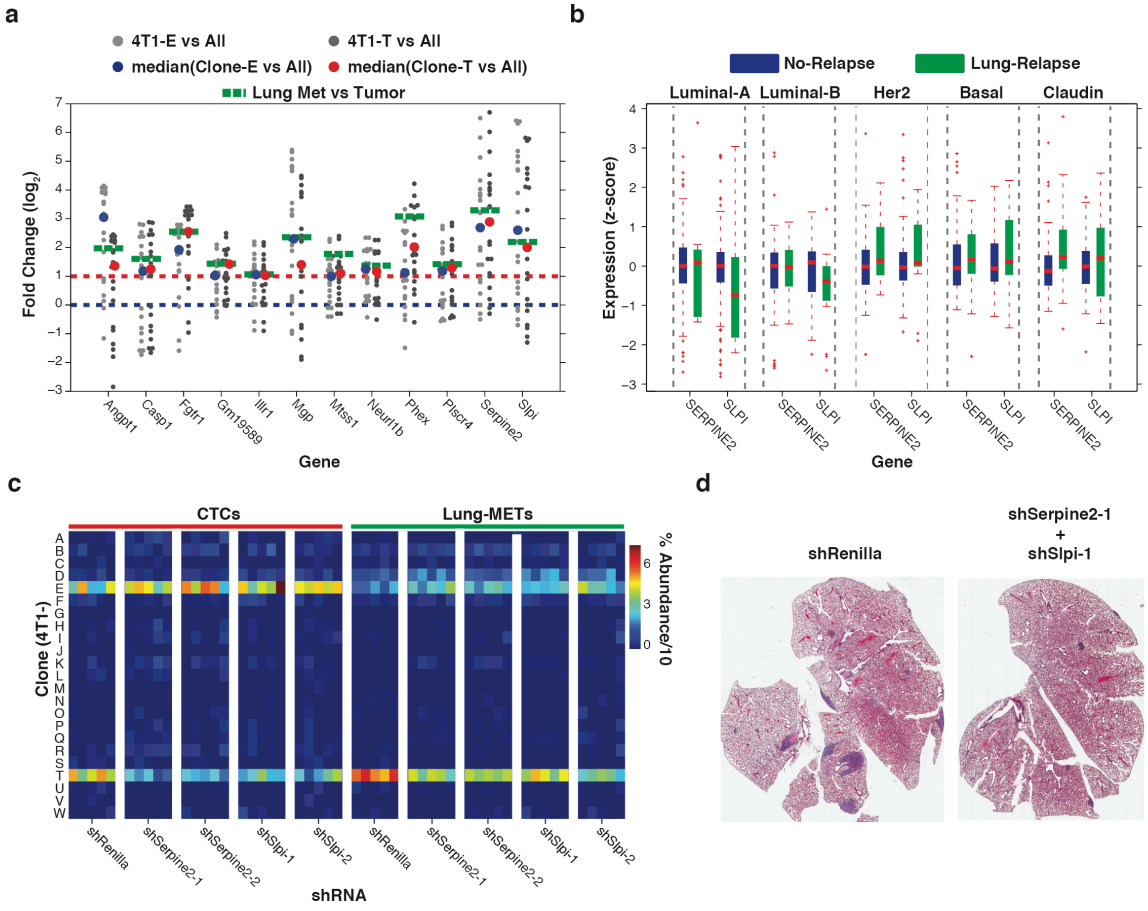


Figure 4.7. Serpine2 and Slpi are regulators of intravasation into the cardiovascular system. (a) Gene expression analysis (via RNA-seq) of each clonal line as well as matched tumor and lung metastases to identify genes upregulated in 4T1-E and -T as compared to the other clones and in lung metastases as compared to primary tumors. All data points above the red dotted line represent at least a 2-fold increase. (b) Analysis of Serpine2 and Slpi human orthologs SERPINE2 and SLPI, respectively in patients with no relapse as compared to patients with relapse in the lung. (c) Relative proportions of clonal lines in the CTCs and lung where 4T1-T has as been infected with non-targeting shRNAs and shRNAs targeting Serpine2 and Slpi. (d) Representative H&E stained lung sections from BALB/c mice that had been orthotopically injected with 4T1 cells that had been infected with a non-targeting shRNA and a vector harboring shRNAs for both Serpine2 and Slpi.

Expression levels of our 12 candidates were additionally examined in human patients, comparing those that did or did not relapse with lung metastases (Harrell et al. 2012). Of the 10 genes with associated patient data, the human orthologs of Serpine2 and Slpi (SERPINE2 and SLPI, respectively) emerged as the most significantly overexpressed in relapsed patients (Fig. 4.7b, p-value < 0.005). The 4T1 cell line is used to model aggressive breast cancer subtypes such as Basal, Her2, and Claudin-Low. Notably, it is precisely these tumor types, and not luminal cancers, that show increased SERPINE2 and SLPI expression in patients that relapse.

Functional validation of Serpine2 and Slpi on lung metastasis

To functionally validate Slpi and Serpine2 as regulators of intravasation, 4T1-T cells were separately infected with two shRNAs for each gene as well as a control shRNA targeting Renilla luciferase (Fig. 4.1c). The separately infected cells were then pooled in equal amounts, orthotopically injected and after 18 days the primary tumors, blood and lungs were removed and their corresponding shRNA abundances quantified via NGS. A comparison of shRNA levels in the tumor vs. the blood and lungs shows a significant depletion of all targeting shRNAs after intravasation (Fig. 4.8a, p-value < 0.01). When the same pool of cells was injected into the left cardiac ventricle, no depletion of targeting shRNAs was observed in the resultant lung metastases as compared to the injection pool, further validating Serpine2 and Slpi as being regulators of intravasation alone (Fig. 4.8b).

To further examine how Serpine2 and Slpi expression shapes this system, the five individually shRNA infected 4T1-T populations were pooled separately with the 22 remaining clonal lines. Each of these five pools, were then orthotopically injected and the clonal analysis as described in Figure 4.5b,c was performed. Figure 4.7c demonstrates that

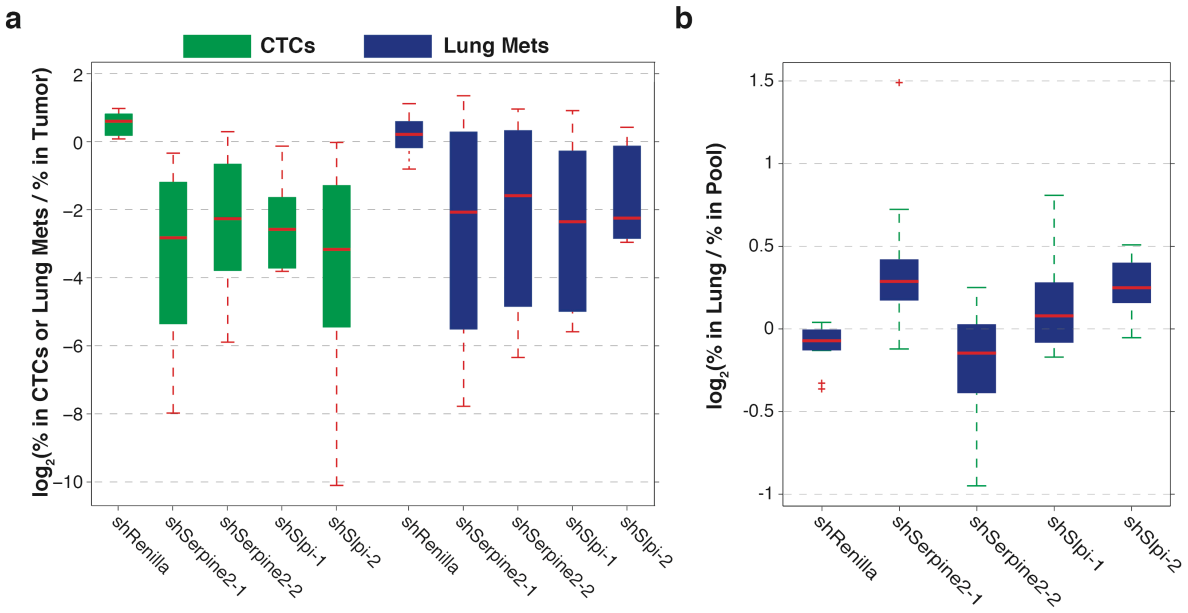


Figure 4.8: Serpine2 and Slpi shRNAs abundances in lung and CTCs after orthotopic and intracardiac injection

(a) Proportions of shRNA targeting Renilla luciferase, Serpine2 and Slpi that had been infected separately into 4T1-T cells. After selection the separately infected cells were pooled in equal amounts with the remaining 22 clonal lines and orthotopically injected into mice. The proportions are in comparison to the primary tumor. (b) Relative proportions of the shRNA infected cells in lung metastases (as compared to the pre-injection pool) 8 days after being injected into the left cardiac ventricle.

depletion of Serpine2 and Slpi in 4T1-T cells resulted in a significant decrease in their representation within the CTCs and lung metastases (p-value < 0.01). In addition, there was a significant reduction in mCherry positive lung nodules (representing shRNA infected 4T1-T cells) in samples corresponding to targeting shRNAs (Fig 4.9a,b, p-value < 0.05).

Finally, to assess how Serpine2 and Slpi act globally across all clones in the 4T1 system in an immune competent setting, 4T1 parental cells were infected with each of the five shRNAs described above as well as with tandem-shRNA vectors harboring combinations of Serpine2 and Slpi targeting shRNAs (Fig 4.1d). These cells were orthotopically injected, and after 18 days secondary lesions in the lung were quantified on H&E stained slides. Lungs corresponding to single gene knockdowns showed significant (~50 %, p-value < 0.01) reduction in metastases, and those corresponding to double-knockdowns showed and even more significant (~60%, p-value < 0.005) reduction in secondary lesions (Fig. 4.7d and 4.9c,d).

Notably, a recent study also implicated Serpins, including Serpine2, in metastasis of breast cancer to the brain (Valiente et al. 2014). These investigators were unable to block metastasis by knockdown of Serpine2. However, in those studies, cells were introduced by intracardiac injection, a procedure that bypasses the requirement for increased Serpine2 expression.

Tumor vasculature and leakiness of clonal cell lines

CD31 is an endothelial cell marker and a major component of their intercellular junctions (Simmons et al. 1990). A qualitative analysis of CD31 stained clonally derived tumor sections revealed a higher number of vessels with focal loss of CD31 staining in 4T1-E and -T tumors, indicating that the vasculature in these tumors might be altered. We wished

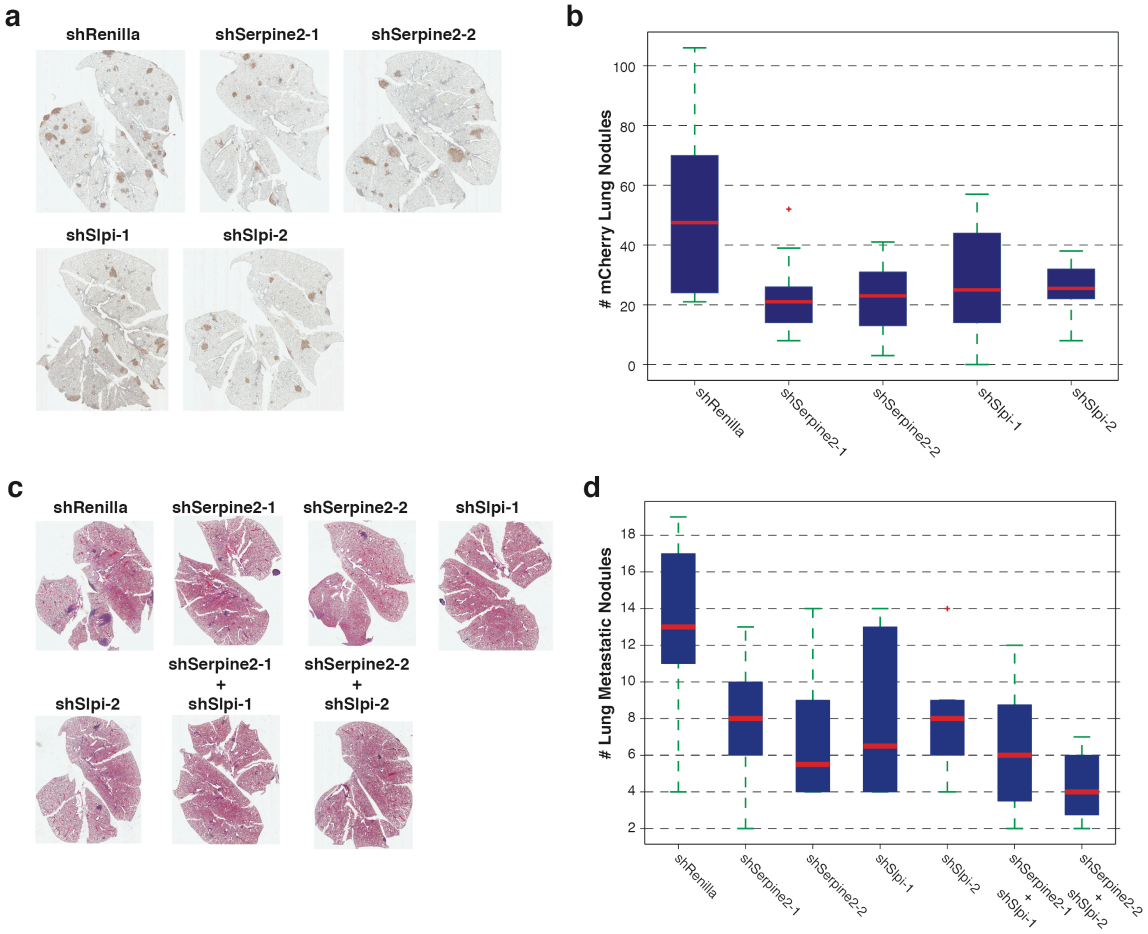


Figure 4.9: Lung metastatic burden of 4T1-T cells or parental 4T1 cells targeted with Serpine2 and/or Slpi shRNAs

(a) Representative mCherry stained sections of the lungs from animals that had been injected with a pool of the 23 clonal lines. In each case, the 4T1-T cells in the pool had been infected with an shRNA targeting Renilla luciferase, Serpine2 or Slpi. The shRNAs were harbored in a vector that constitutively expressed mCherry. (b) Quantification of all mCherry positive metastatic lung nodules in all mice. (c) Representative H&E stained lung sections from animals that had been orthotopically injected with parental 4T1 cells. In each case the cells were infected with a retroviral vector constitutively expressing an shRNA targeting Renilla luciferase, Serpine2 or Slpi or a tandem vector targeting both Serpine2 and Slpi. (d) Quantification of the number of metastatic nodules in all H&E stained sections described above.

to quantify these observations and we reasoned that in tumors where vasculature is altered, a higher degree of vascular leakiness should be observed. Thus, we developed and qualified an analysis pipeline to assign a leakiness index to tumors using a high molecular weight dextran (see Methods). Using this pipeline we determined that the leakiness of 4T1-T derived tumors was significantly higher than in 4T1-L derived tumors (Fig 4.10a,b and 4.11a). Silencing of Serpine2 or Slpi resulted in a reduction in leakiness, implicating these proteins as drivers of this vascular phenotype (Fig. 4.11b, p-value < 0.05).

We hypothesized that direct contact of tumor cells with blood, due to tumor leakiness, might result in tumor cells being passively drawn into the bloodstream to form CTCs. To test this hypothesis the clonal lines were pooled and orthotopically injected into three mice cohorts. One cohort received regular drinking water while the other two received drinking water with 10 mg/L warfarin (mimicking tumor leakiness because of the inability to coagulate blood properly). At this dose all animals survive, prothrombin activator fragments F1 + F2 are reduced by > 80% in the plasma, and a significant increase in tumor leakiness is observed (Fig. 4.12a,b,c). One treated cohort began treatment 1 day prior to injection and the other 10 days post-injection. In addition, cells were injected through the left cardiac ventricle of three additional cohorts, and these received the same warfarin dose-regimes listed above.

Previous work has shown that, due to a reduction in clotting in the lung, when cells are injected directly into the blood, a decrease in metastases is observed when warfarin is administered (McCulloch and George 1987; Bobek and Kovarík 2004; Mousa 2006). This result was recapitulated in mice that were injected via the left cardiac ventricle (Fig. 4.12d, p-value < 0.001). In contrast, when intravasation was made a requirement (via orthotopic injection), administration of warfarin significantly increased the lung-metastatic burden (Fig.

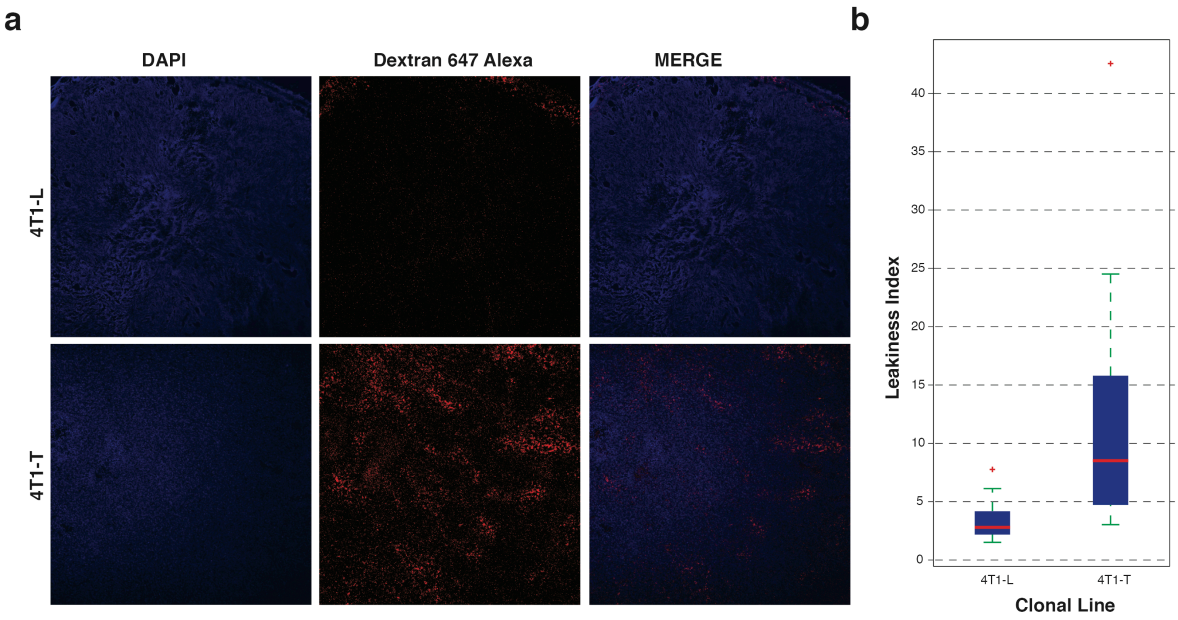


Figure 4.10: Vascular leakiness in 4T1-L and 4T1-T primary tumors
(a) DAPI and Dextran, Alexa 647 (+ MERGED) stained tumor sections from orthotopic tumors derived from 4T1-L and –T cells. (b) Leakiness Index of primary tumors resulting from orthotopic injection of 4T1-L or –T cells.

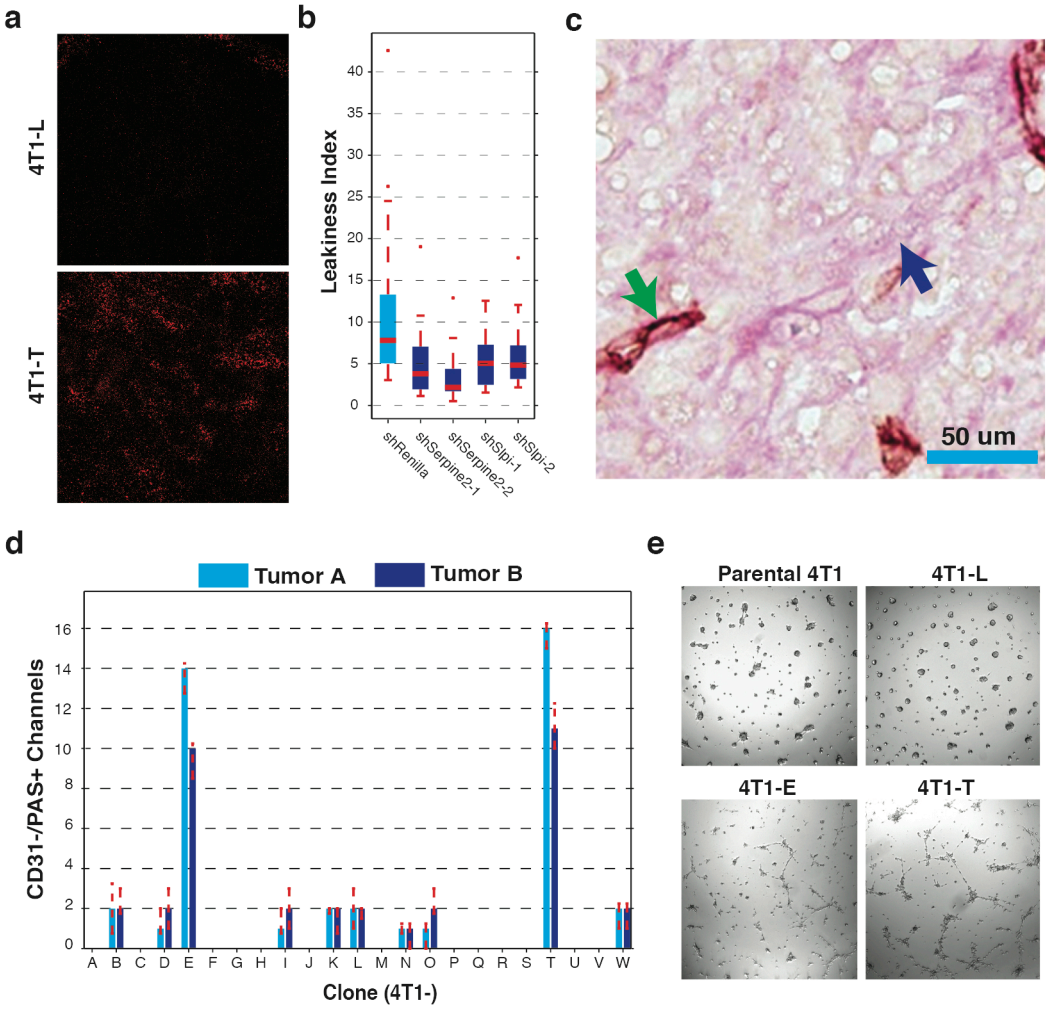


Figure 4.11. Vascular mimicry drives metastatic progression
(a) Vascular leakage in sections taken from primary tumors derived from clonal lines 4T1-L and –T that were intravenously injected with high molecular weight Dextran, Alexa 647. (b) The leakiness index (see Methods) of primary tumors derived from 4T1-T cells that have been infected with either a non-targeting shRNA or an shRNA targeting Serpine2 or Slpi. (c) CD31 + PAS stained section of a 4T1-T derived primary tumor. Green arrow indicates a CD31 positive blood vessel. Blue arrow indicates a CD31 negative, PAS positive channel. (d) Quantification of CD31-/PAS+ channels in clonally derived tumors from each clonal line. (e) Images of Parental 4T1, 4T1-L, 4T1-E and 4T1-T cells grown on matrigel.

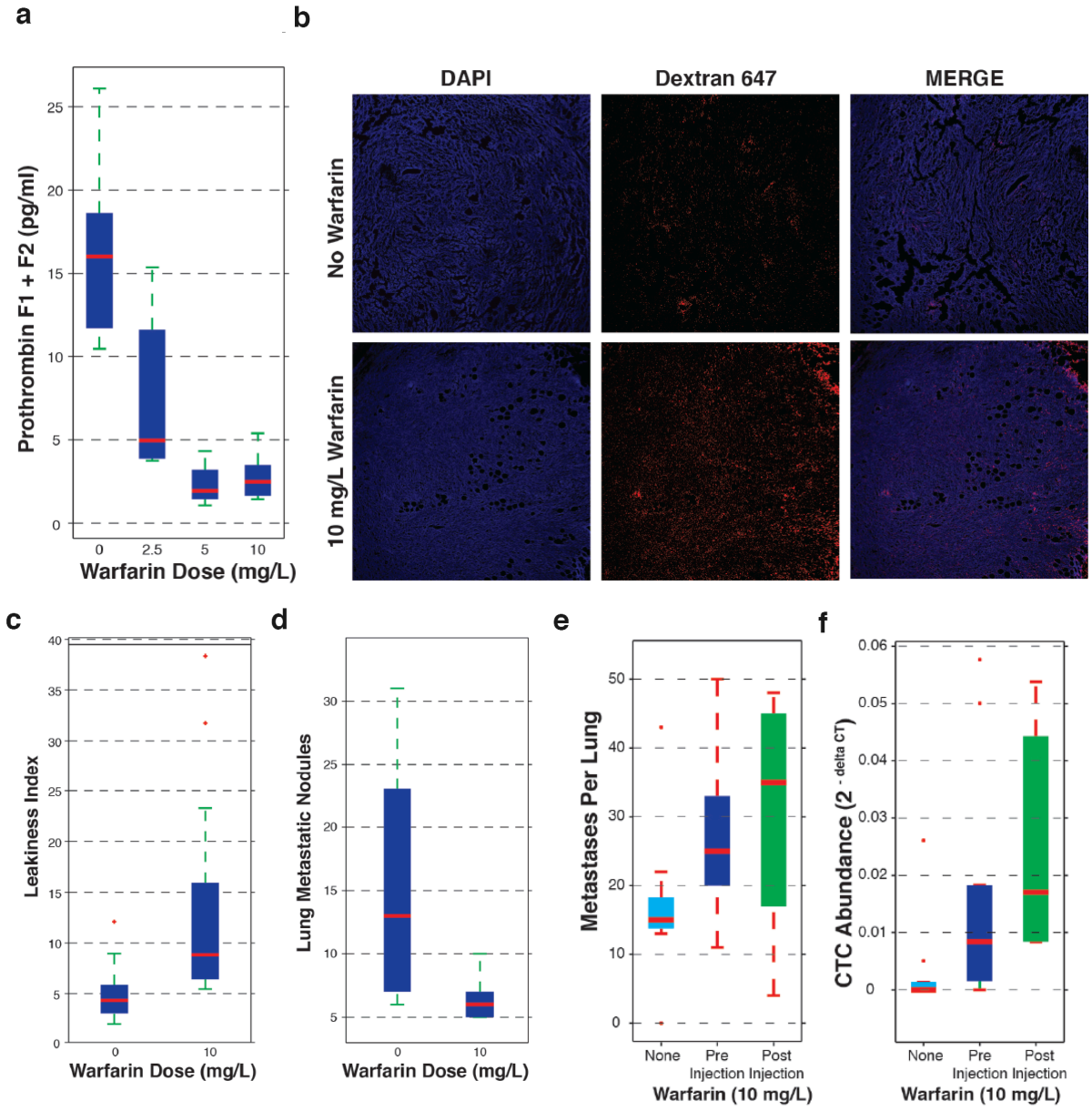


Figure 4.12: Effect of Warfarin on metastatic progression
(a) Reduced levels of cleaved Prothrombin F1 + F2 (active component of blood coagulation) upon administration of Warfarin in drinking water. (b) Vascular leakage in sections taken from primary tumors derived from all 23 clonal lines that were intravenously injected with high molecular weight Dextran, Alexa Fluor 647. Animals were administered drinking water either with or without Warfarin. (c) Leakiness Index of primary tumors resulting from orthotopic injections of all 23 clonal lines. (d) Counts of lung metastatic nodules from animals there were injected intravenously with all 23 clonal lines and were treated either with or without Warfarin. (e) Counts of lung metastatic nodules from animals there were injected orthotopically with all 23 clonal lines and were treated either with Warfarin pre- or post-injection or without Warfarin. (f) CTC abundance (measured by qPCR) from animals that were injected orthotopically with all 23 clonal lines and were treated either with Warfarin pre- or post-injection or without Warfarin.

4.12e). Further, via qPCR of whole blood for sequences within the barcode-harboring vector, we determined that this was the result of an increased CTC load (Fig. 4.12f). Finally, an analysis of barcodes showed that 4T1-E and -T remained pervasive in the blood of the warfarin treated animals, indicating that they were preferentially drawn into the blood stream by the treatment (Fig 4.13). This was taken as evidence that these clones are likely in close proximity to sites of vascular leakiness in the primary tumor.

Vascular mimicry of clonal cell lines

The characteristics of 4T1-E and -T derived tumors, namely focal loss of CD31 in vessels and a high degree of tumor leakiness, have been associated with a recently described phenomenon termed vascular mimicry (VM), where tumor cells differentiate into endothelial-like cells and form extracellular matrix (ECM) rich tubular structures to carry blood from the vasculature to hypoxic regions of the tumor (Fig. 4.11c) (Maniotis et al. 1999; Folberg et al. 2000; Hendrix et al. 2003; Francescone et al. 2011). We hypothesized that the ability of 4T1-E and –T cells to intravasate was the result of a heightened capacity for VM. This would place the cells in direct contact with the blood, which we have demonstrated above results in increased intravasation. To test this hypothesis we assessed CD31 + Periodic acid–Schiff (PAS) stained sections of clonally derived tumors from all lines for characteristics consistent with VM (see Methods). This analysis revealed a striking increase in these characteristics for 4T1-E and –T tumors as compared to the others (Fig. 4.11.d and 4.14a-e, p-value < 0.001).

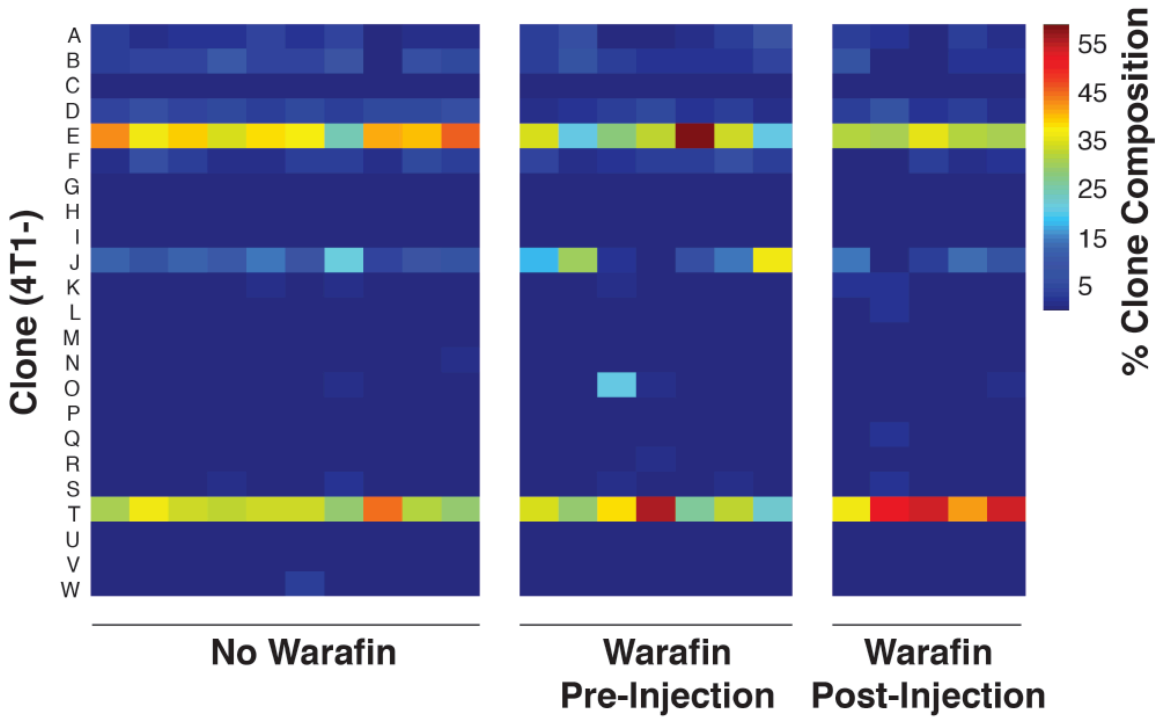


Figure 4.13: Clonal barcode composition upon Warfarin treatment
Barcode representation of animals that were orthtopically injection with a mix of all 23 clonal lines and were treated either with Warfarin pre- or post-injection or without Warfarin.

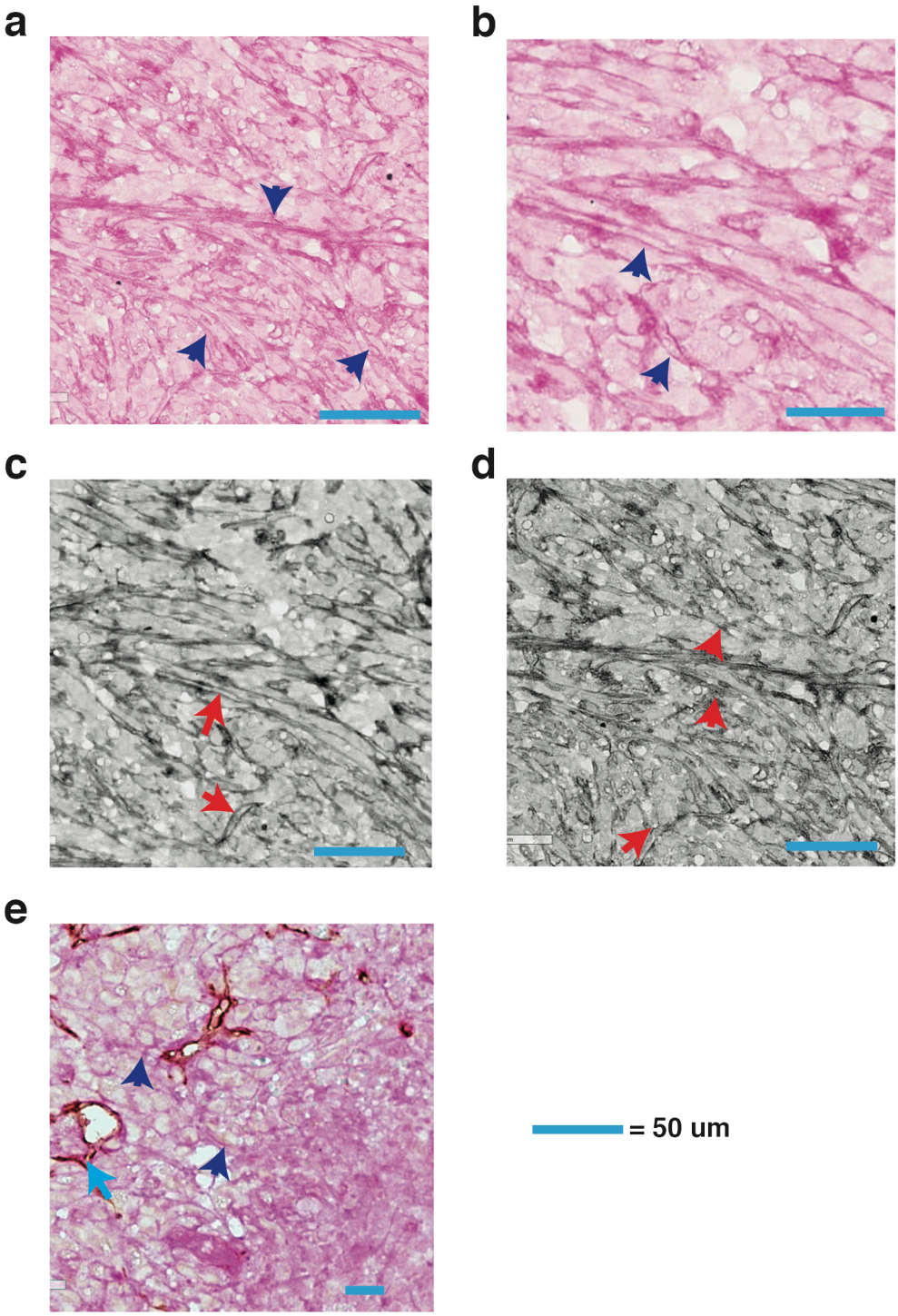


Figure 4.14: Vascular mimicry in 4T1-E and 4T1-T primary tumors
(a) Periodic acid-Schiff (PAS) staining of 4T1-E derived primary tumors. Arrows indicate networks of open channels bordered by thin PAS positive extracellular matrix. (b) PAS staining of a 4T1-T derived primary tumor. Arrows indicate the same PAS positive channels as described in (a). (c) Black and white image of PAS stained 4T1-E derived primary tumors. Arrows indicate anastomosing vascular channels. (d) Black and white image of PAS stained 4T1-T derived primary tumors. Arrows indicate anastomosing vascular channels. e. CD31 + PAS stained 4T1-T derived primary tumor. Light blue arrows indicate vessels with interrupted CD31 staining. Dark blue arrows indicate CD31 negative PAS positive

When human umbilical vein endothelial cells (HUVEC) are grown on matrigel they form a rich tubular network, just as do tumor cells capable of vascular mimicry (Fig. 4.15a) (Maniotis et al. 1999; Francescone et al. 2011). We decided to apply this method to assess 4T1-E and T's capacity for VM. When parental 4T1 cells were plated there was weak evidence for VM and when the non-intravasating clone 4T1-L was plated in matrigel no evidence for VM was observed (Fig. 4.11e). In contrast, when 4T1-E or 4T1-T cells were plated, strong evidence for VM was observed, validating the analysis performed in Figure 4.11d.

To determine if Serpine2 or Slpi promote VM, we plated the 5 shRNA infected 4T1-T cell lines on matrigel. The cells infected with the non-targeting shRNA formed strong tubular networks. In contrast almost all evidence for VM was abolished in cells infected with the targeting shRNAs (Fig. 4.16a). Taken together with the data presented in Figure 4.5c,d this indicates that clone 4T1-E and T's proclivity for intravasation is the result of Serpine2 and Slpi driven VM.

Next, to determine if an increased expression of Serpine2 and/or Slpi might promote VM and intravasation in non-intravasating clones we overexpressed Serpine2 and Slpi in 4T1-B, F, N and S cells. We determined that in all cases, when Serpine2 or Slpi were over expressed, a significant increase in VM resulted (Fig. 4.16b, p-value < 0.05). In addition, to test if this same phenomenon could be observed in human cells, we also overexpressed the two splicing variants of SERPINE2 as well as SLPI in three human breast cancer lines: MDA-MB-231 (Basal), MDA-MBA-436 (Basal) and T47D (Luminal). We hypothesized that since expression of SERPINE2 and SLPI is only predictive of relapse in aggressive tumors (Fig. 4.7b) that only the Basal cell lines would respond to over expression. True to our hypothesis, in both Basal cell lines a significant increase in VM was observed when over expression was

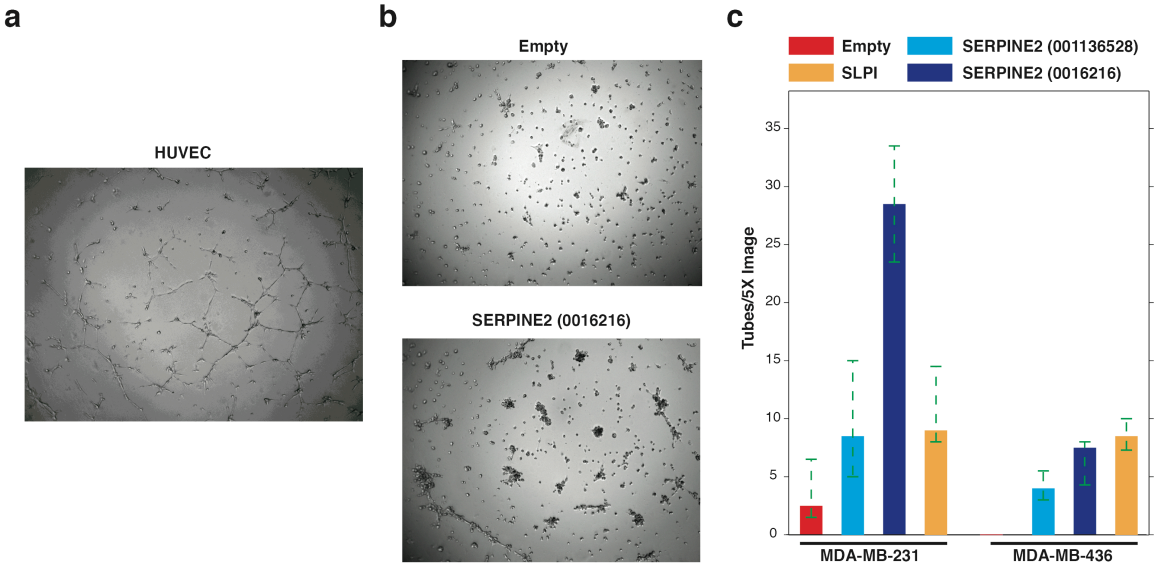


Figure 4.15: Tube formation assay of human breast cancer cells upon overexpression of SERPINE2 and SLPI

(a) Human umbilical vein endothelial cells (HUVEC) plated on matrigel to assess tube formation ability. (b) MDA-MB-231 cells infected with an empty overexpression vector plated on matrigel (upper panel) and MDA-MB-231 cells overexpressing SERPINE2 (transcript NM_0016216) plated on matrigel (lower panel). (c) Full quantification of tubular structures in basal cell lines MDA-MB-231 and MDA-MB-436 when infected with an empty overexpression vector or when infected with vectors for overexpressing SERPINE2 (transcript NM_0016216), SERPINE2 (transcript NM_001136528) or SLPI.

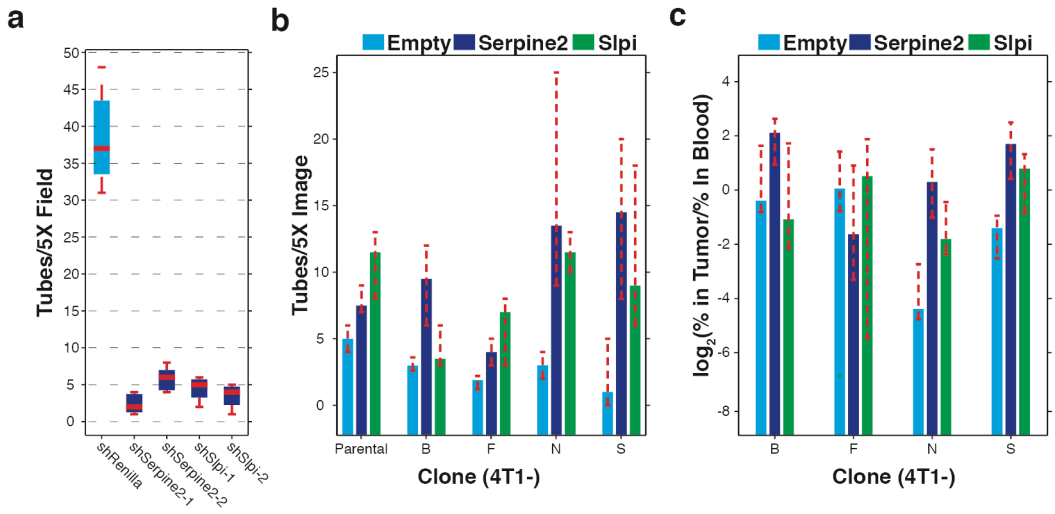


Figure 4.16. Serpine2 and Slpi2 drive vascular mimicry
(a) Number of tubular structures identified per 5X microscopic field when 4T1-T cells had been infected with a non-targeting shRNA or shRNAs targeting Serpine2 or Slpi and grown on matrigel. (b) Number of tubular structures identified per 5X microscopic image when non-intravasating 4T1 clonal lines have been infected with an empty retroviral vector or vectors for overexpression of Serpine2 or Slpi. (c) Relative CTC proportions of the cells described in G after they had been pooled with the remaining clonal lines and orthotopically injected.

induced, however no evidence for VM was observed in the Luminal line (Fig. 4.15b,c, p-value < 0.05).

Finally, we wished to determine if the induced increase in VM in 4T1-B, F, N and S cells resulted in their having an increased proclivity to intravasate. The clones over-expressing Slpi, Serpine2 as well as those harboring an empty vector were separately pooled with the remaining 19 clonal lines (with the exception of 4T1-E, to more easily detect increases in the overexpressing clones intravasation). The pools were then orthotopically injected and, after 18 days, the clone profiles of the tumors and blood were profiled. True to the hypothesis that VM drives intravastaion, with the exception of 4T1-F, an increase in VM resulting from overexpression of Serpine2 or Slpi resulted in an increased capacity for intravasation (Fig. 4.16c, p-value < 0.05).

General conclusions

We have described a mouse model of breast tumor heterogeneity, which allows us to probe the molecular basis of stable differences in the ability of clonal populations to contribute to various aspects of disease. In this initial study, we focused on those characteristics distinguishing clones that efficiently contribute to CTCs, a subset of which also form metastases. The ability to form CTCs, and ultimately metastases, is closely linked to the capacity for vascular mimicry. Tumor cell lined vasculature has shown a strong clinical correlation with advanced stage disease and poor clinical outcome (Cao et al. 2013). In our model, vascular mimicry is driven by increased expression of two secreted proteins, Serpine2 and Slpi. Very little is currently known of the molecular determinants that enable vascular mimicry, and Serpine2 and Slpi are among the first validated drivers of this process. These proteins are both serine protease inhibitors, and the precise mechanism by which they

might confer vascular mimicry is unclear, though Serpine2 promotes ECM production and VM channels are ECM-rich (Hendrix et al. 2003; Buchholz et al. 2003). Both Serpine2 are Slpi are also anticoagulants (Bouton et al. 2012; Masuda et al. 1995). Thus, they may also play a role in keeping open an interface between a tubule created by tumor cells and the tumor vasculature, allowing passage of both red blood cells into the tumor and cancer cells into the bloodstream. Notably, these dual roles of vascular mimicry link promotion of metastasis to satisfaction of a need of the primary tumor, essentially providing a way in which metastatic potential might be positively selected during the process of primary tumor formation.

Others are making great strides in growing patient material in various culture systems. It is our hope that by applying these technologies we will be able to apply this method to patient material to identify the most likely cells to metastasize so that treatment strategies can be adjusted accordingly.

Methods

Cell culture

The mouse mammary tumor cell line 4T1 (ATCC) and any derived clonal cell line were cultured in DMEM high glucose (Life Technologies) supplemented with 5% fetal bovine serum (Thermo Scientific), 5% fetal calf serum (Thermo Scientific), non-essential amino acids (Life Technologies) and penicillin streptomycin (Life Technologies). Human breast tumor cell lines MDA-MB-231, MDA-MB-436 and T47D (ATCC) were cultured in DMEM high glucose supplemented with 10% fetal bovine serum, non-essential amino acids and penicillin streptomycin. Human umbilical vein endothelial cells (HUVEC) (Lonza) were cultured in EBM-2 media with the EGM-2 Bulletkit (Lonza). HUVEC cells were used within 3 passages. Platinum-A (Cell BioLabs) and 239-FT (Life Technologies) packaging cell lines were cultured in DMEM high glucose supplemented with 10% fetal bovine serum and penicillin streptomycin.

Virus production

All retroviral vectors were packaged using Platinum-A packaging cells. The lentiviral barcode library was packaged using 293-FT lentivirus packaging cells. Cells were plated on 15 cm adherent tissue culture plates (Corning) ~5 hours before transfection at a confluency of ~70%. A transfection mixture was prepared with viral vector (75 μg), VSV-G (7.5 μg), 2M calcium chloride (187.5 μl) (Sigma-Aldrich) and, when transfecting shRNA harboring vectors, 20 nM siRNAs targeting Pasha (200 μl) (Qiagen). The mixture was brought to 1.5 ml with H_20 and then added drop-wise to the same amount of 2x HBS while being bubbled. One liter of 2X HBS was prepared with 280 mM NaCl, 50 mM HEPES, 1.5 mM Na2HPO4, 12 mM Dextrose and 10 mM KCl (Sigma-Aldrich), then adjusted to a pH of

7.02. After the transfection mixture was added to the HBS, vigorous bubbling continued for 30 - 60 seconds. After letting the resultant mixture stand for 15 minutes, it was added to the packaging cells along with 100 mM chloroquin (7.5 ul) (Sigma-Aldrich). After 14 hours, media was replaced. 30 hours after media change, virus was collected and filtered through a 0.45 μm filter (EMD Millipore) and stored at 4°C.

Establishment of clonal cell lines

30 million 4T1 cells were infected with the lentiviral barcode library (Fig. 4.2a) at a multiplicity of infection (MOI) of 0.3. Single cells were sorted using the FACSAria IIU cell sorter (BD Biosciences) into 96 well plates. Clonal cell lines were minimally expanded and frozen down. For all experiments, each clonal cell line was used within 7 passages. The barcode of each individual clonal cell line was determined by Sanger sequencing. Forward primer Comp-Mir5: 5- CAG AAT CGT TGC CTG CAC ATC TTG GAA AC -3 and reverse primer Comp-WPRE: 5- ATC CAG AGGTTG ATT GTT CCA GAC GCG T -3.

Clonal cell line proliferation rates

Proliferation assays were performed by counting viable cells over 72 hours. 1 x 10^5 cells were plated in duplicates and were counted using the MACSQuant Analyzer (Miltenyi Biotec).

Chromosomal integration site

Genomic DNA from each clone was isolated using the QIAamp DNA blood mini kit (Qiagen). Choromosmal integration sites were determined using the Retrovirus Integration Site Analysis Kit (Clontech).

Animal studies

All mouse experiments were approved by the Cold Spring Harbor Animal Care and Use Committee. All orthotopic injections were performed using 1 x 10^5 mouse mammary tumor cells resuspended in 20 ul of a 1:1 mix of PBS and growth factor reduced Matrigel (BD Biosciences). Injections were into mammary gland #4 of 8 week-old female NOD scid gamma mice (JAX). For intracardiac injections, 1 x 10^5 mouse mammary tumor cells were resuspended in 200 ul of PBS and injected into the left cardiac ventricle. For tail vein injections, 5 x 10^5 mouse mammary tumor cells were resuspended in 100 ul of PBS.

For clonal pooling experiments, clonal cell lines were counted in duplicate using the MACSQuant Analyzer (Miltenyi Biotec). Equal numbers of cells were pooled together for injection. A pre-injection pool was collected to validate equal representation of each clone before injection. Tumor, lung, brain, liver, brachial lymph node and blood were harvested from mice for further processing. Primary tumor volume was measured using the formula $V = 1/2(LxW^2)$, where L is length and W is width of the primary tumor. Warfarin (10 mg/L, Sigma-Aldrich) was administered with drinking water and changed every 3 days.

Quantification of lung metastatic burden

The lung metastatic burden of individual clones injected into the mammary gland was evaluated in five-micron sections stained with a standard H&E protocol. Four mice were evaluated per clone. Quantification was performed using ImageJ Software (NIH) converting images to 8-bit. Upper and lower thresholds for each image were adjusted to determine total lung area and adjusted again to determine the metastatic area. Both values were used to obtain relative metastatic areas.

For other experiments, the lung metastatic burden was evaluated by counting the number of metastatic nodules in the lung. For this, five-micron sections were stained with a standard H&E protocol. For all experiments, ten mice were evaluated per condition.

Barcode and shRNA analysis

Genomic DNA was isolated using phenol chloroform extraction for all tissues except blood. Genomic DNA for blood was isolated using the QIAamp DNA Blood Mini Kit (Qiagen).

The barcodes of the retroviral library (Fig. 4.2a) were amplified using a one-step PCR protocol. For each sample, 96 individual PCR reactions of 200 ng of genomic DNA were carried out using KOD Polymerase (EMD Millipore). Forward primer: 5- AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T -3 and reverse primer: 5- CAA GCA GAA GAC GGC ATA CGA GAT NNN NNN GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC -3. The reverse primer contained a barcode (NNNNNN) that enabled multiplexing with standard Illumina Truseq chemistry and software. The PCR was carried out for 30 cycles and PCR products were purified using a PCR purification kit (Qiagen). PCR products were size selected on an E-gel SizeSelect 2% agarose gel (Life Technologies), and sequenced on the Illumina HiSeq sequencer generating 22 nt single-end (SE) reads.

The barcodes of the lentiviral library (Fig. 4.2b) were amplified using a two-step PCR protocol. For each sample, eight individual PCR reactions of 200 ng of genomic DNA were carried out using KOD Polymerase (EMD Millipore). Forward primer 1: 5- GTG ACT GGA

GTT CAG ACG TGT GCT CTT CCG ATC TCA GAA TCG TTG CCT GCA CAT CTT GGA AAC -3 and reverse primer 1: 5- ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT ATC CAG AGG TTG ATT GTT CCA GAC GCG T -3. The first PCR was carried out for 25 cycles. PCR products were purified using the PCR purification kit (Qiagen). The second PCR was performed using 500 ng of PCR product from the first PCR. Forward primer 2: 5- AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T -3 and reverse primer 2: 5- CAA GCA GAA GAC GGC ATA CGA GAT NNN NNN GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC -3. The reverse primer contained a barcode (NNNNNN) that enabled multiplexing with standard Illumina Truseq chemistry and software. The second PCR was carried out for 25 cycles and PCR products were again purified using the PCR purification kit (Qiagen). PCR products were size selected on an E-gel SizeSelect 2% agarose gel (Life Technologies), and sequenced on the Illumina HiSeq sequencer generating 22 nt single-end (SE) reads.

For Figure 4.1 and 4.5d, the vector library was sequenced at high depth. For each experiment the corresponding fastq file was aligned to the vector library with the Bowtie software, allowing 3 mismatches. Each experimental read was then assigned to the most abundant vector sequence that it mapped to. For Figure 4.1 only sequences that were present with a count greater than or equal to 5 in all tumors were analyzed.

For Figure 4.3b,c abundant clones were identified by fitting a mixed Gaussian model to the summed distributions described in Figure 4.3a. Abundant clones were those that then subsequently clustered into the Gaussian with the larger mean.

The shRNAs were amplified using the same two-step PCR protocol as described above for the lentiviral barcode library. Forward primer 1: 5- CAG AAT CGT TGC CTG CAC ATC TTG GAA AC -3 and reverse primer 1: 5- CTG CTA AAG CGC ATG CTC CAG ACT GC -3. Forward primer 2: 5- AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT AGC CTG CGC ACG TAG TGA AGC CAC AGA TGT A -3 and reverse primer 2: 5- CAA GCA GAA GAC GGC ATA CGA GAT NNN NNN GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TCT GCT AAA GCG CAT GCT CCA GAC TGC -3. The reverse primer contained a barcode (NNNNNN) that enabled multiplexing.

RNA-Seq library preparation and analysis

Total RNA was purified and DNAse treated using the Qiagen RNeasy Mini Kit. RNA integrity (RNA Integrity score > 9) and quantity was measured on an Agilent Bioanalyzer (RNA Nano kit). The NuGEN Ovation RNA-Seq V2 protocol was carried out on 100 ng of total RNA. cDNA was fragmented using the Covaris LE220 sonicator according to the manufacturer's instruction to yield a target fragment size of 200 bp. The fragmented cDNA was subsequently processed using the NuGEN Ovation Ultralow DR Multiplex System. Each sample was sequenced on the Illumina HiSeq sequencer generating 76 nt single-end (SE) reads.

Reads were aligned to the mm10 genome using the Bowtie-2 alignment tool. Mapped reads were then assigned to genes using HTSeq-count (using the latest version of RefSeq.gtf file for gene coordinates). Resultant counts were then normalized and compared using DESeq. For a gene to be considered over-expressed it had to show an at least 2-fold change with FDR < 0.05.

Isolation of matched tumor and lung tissue

Tumor and lung tissue were harvested, minced and treated in DMEM high glucose containing 1x collagenase/hyaluronidase buffer (StemCell) and 10 U DNAse I (Sigma) for 1 hour at 37 C. Cells were washed in HBSS (Life Technologies) two times and then re-suspended in 4T1 cell culture media containing 60mM 6-Thioguanine. Cells were passaged for 5 days until all stromal cells died.

shRNA knock-down and cDNA overexpression

Mouse and human cell lines were transduced with amphotropically packaged retroviruses. MDA-MBA-231 and T47D cells were selected with 1500 mg/ml G418 and MDA-MD-436 cells were selected with 1000 mg/ml G418 for 1 week. All mouse clonal cells lines were selected with 1000 mg/ml G418 for 1 week, the parental 4T1 cell line was selected with 600 mg/ml G418 for 1 week.

qRT-PCR

Total RNA was purified and DNAse treated using the RNeasy Mini Kit (Qiagen). Synthesis of cDNA was performed using SuperScript III Reverse Transcriptase (Sigma). Quantitative PCR analysis was performed on the Eppendorf Mastercycler ep realplex. All signals were quantified using the ΔCt method and were normalized to the levels of Gapdh.

qRT-PCR primers

mouse Slpi (Exon 1-2):

5'-GAC TGT GGA AGG AGG CAA A-3'

5'-GGC ATT GTG GCT TCT CAA G-3'

mouse Slpi (Exon 3-4):

5'-CAG TGT GAC GGC AAA TAC AAG-3'

5'-GCC AAT GTC AGG GAT CAG G-3'

mouse Serpine2 (Exon 3-4):

5'-TCT GCC TCT GAG TCC ATC A-3'

5'-AAC CGA GAC TTC CAC AAA CC-3'

mouse Serpine2 (Exon 5-6):

5'-TCA TCC CTC ACA TCA CTA CCA-3'

5'-CTT TCA GTG GCT CCT TCA GAT-3'

mouse Gapdh (Exon 2-3):

5'-AAT GGT GAA GGT CGG TGT G-3'

5'-GTG GAG TCA TAC TGG AAC ATG TAG-3'

human SLPI (Exon 1-2):

5'-TGT GGA AGG CTC TGG AAA G-3'

5'-TGG CAC TCA GGT TTC TTG TAT C-3'

human SERPINE2 (Exon 5-6):

5'-GCC ATG GTG ATG AGA TAC GG-3'

5'-GCA CTT CAA TTT CAG AGG CAT-3'

human GAPDH (Exon 2-3):

5'-ACA TCG CTC AGA CAC CAT G-3'

5'-TGT AGT TGA GGT CAA TGA AGG G-3'

Figure 4.17 shows knock-down efficiencies of Serpine2 and Slpi shRNAs (mouse) and fold overexpression of Serpine2 and Slpi (mouse and human)

Prothrombin Fragment 1+2 (F1+2) ELISA

Blood was collected from animals that were treated with warfarin by cardiac heart puncture using 3.8% sodium citrate as an anticoagulant. Samples were centrifuged at 1000g for 15 min at 4 C. Blood plasma was isolated and stored at -80 C for further processing. Plasma samples were analyzed for Prothrombin Fragment 1+2 using the Mouse Prothrombin Fragment 1+2 (F1+2) ELISA kit (Kamiya Biomedical Company) according to manufacturer's instructions. Each sample was measured in duplicates and each condition had four to five animals.

mCherry analysis

For immunoflourescence, five-micron sections of paraffin-embedded lungs were deparaffinized in xylene, rehydrated in an alcohol series and immersed in distilled water. The sections were then treated with high-temperature antigen retrieval in citrate buffer (pH 6), blocked with MaxBlock (ActiveMotif) for 1 hour at 37 C and incubated with primary antibody RFP Antibody Pre-adsorbed (1:100) (600-401-379, Rockland) overnight at 4 C. After washing, the slides were incubated for 1 h with Alexa 647 donkey anti-rabbit IgG (1:500) (A-31573, Life Technologies) and counterstained with DAPI (1 mg/ml) (Sigma-

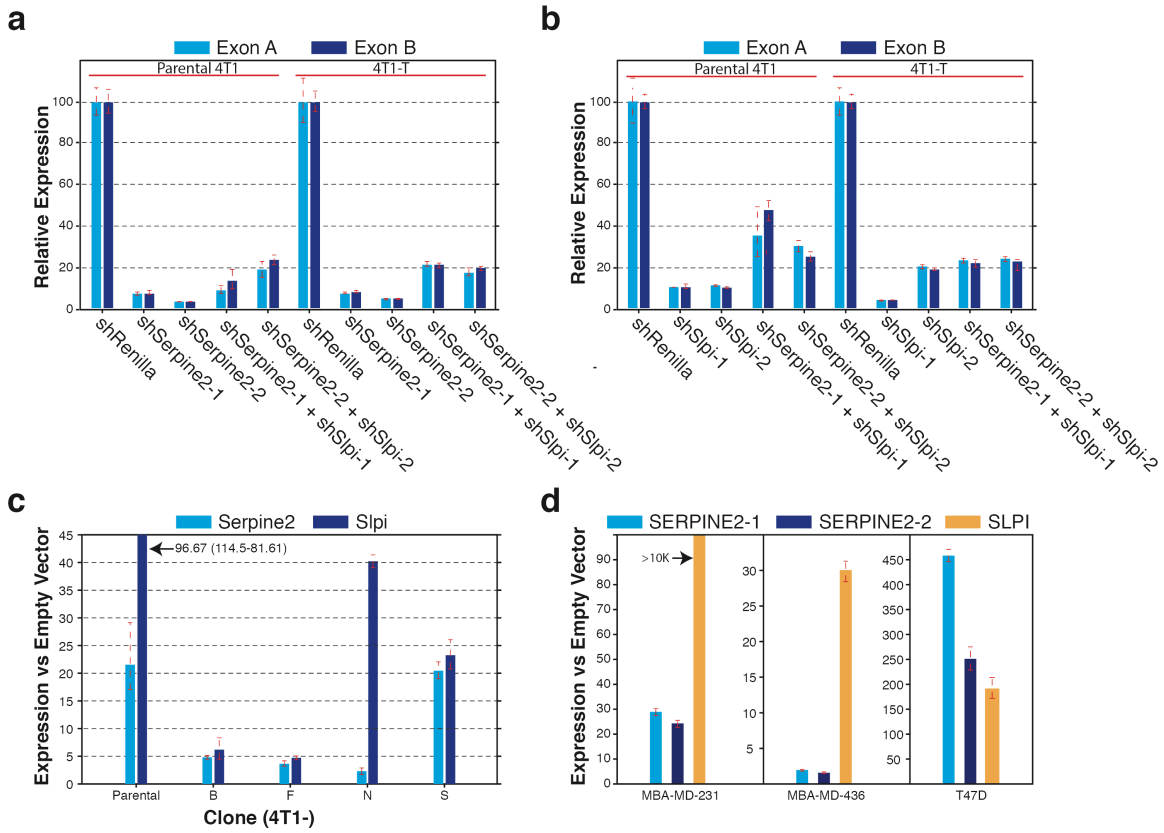


Figure 4.17: Knock-down efficiency of Serpine2 and Slpi shRNAs and fold overexpression of SERPINE2 and SLPI cDNA

(a) Knockdown efficiency of Serpine2 (as measured by qPCR) in parental 4T1 and 4T1-T cells after infection with each of the constructs harboring an shRNA targeting Serpine2. (b) Knockdown efficiency of Slpi (as measured by qPCR) in parental 4T1 and 4T1-T cells after infection with each of the constructs harboring an shRNA targeting Slpi. (c) Fold overexpression of Serpine2 and Slpi in parental 4T1 as wells as 4T1-B, F, N and S cells after infection with overexpression vectors harboring Serpine2 and Slpi cDNA (as calculated relative to cells infected with empty vectors). (d) Fold overexpression of SERPINE2 transcripts 1 (transcript NM_006216) and 2 (transcript NM_001136528) as well as SLPI in MBA-MD-231, MBA-MB-436 and T47D cells after infection with overexpression vectors harboring SERPINE2 transcripts 1 and 2 as well as SLPI cDNA (as calculated relative to cells infected with empty vectors). The high relative expression of SLPI in MDA-MB-231 cells infected with the SLPI overexpression vector is due to near zero SLPI levels in the parental cell line.

Aldrich) for 5 minutes. Slides were mounted with ProLong Gold antifade reagent (Life Technologies) and cover-slipped. Sections were scanned on the AperioFL Fluorescence Slide Scanner (Aperio).

For immunohistochemistry, the same initial steps were followed as for immunofluorescence. The slides were then blocked with 2.5% ready-to-use normal horse serum from ImmPRESS Anti-Rabbit Ig (peroxidase) Polymer Detection Kit (MP-7401, Vector Laboratories) for 1 hour and then incubated with primary antibody RFP Antibody Pre-adsorbed (1:200) (600-401-379, Rockland) overnight at 4 C. After washing, the slides were incubated with secondary antibody from the previous kit for 30 minutes, rinsed, and developed with chromogen ImmPACT DAB Peroxidase Substrate for staining (SK-4105, Vector Laboratories) until the desired intensity was achieved. Slides were counterstained with hematoxylin and coverslipped. Sections were then scanned on the Aperio Light Field Slide Scanner (Aperio) for further quantification.

Omission of the primary antibody was used as a negative control in both cases.

CD31 and Periodic acid-Schiff analysis

Five-micron sections of paraffin-embedded lungs were de-paraffinized in xylene, rehydrated in an alcohol series and immersed in distilled water. The sections were then treated with high-temperature antigen retrieval in citrate buffer (pH 6), blocked with 2.5% ready-to-use normal horse serum from ImmPRESS Anti-Rabbit Ig (peroxidase) Polymer Detection Kit (MP-7401, Vector Laboratories) for 1 hour and incubated with primary antibody against CD31 (1:400) (28364, Abcam) overnight at 4 C. After washing, the slides were incubated with secondary antibody from the previous kit for 30 minutes, rinsed, and

developed with chromogen ImmPACT DAB Peroxidase Substrate for staining (SK-4105, Vector Laboratories) until the desired intensity was achieved. Slides were then stained with Periodic Acid-Schiff (PAS) Kit (Sigma) according to manufacturer's instructions. Sections were then scanned on the Aperio Light Field Slide Scanner (Aperio) for further analysis. Omission of the primary antibody was used as a negative control.

Two tumors from each clone and the parental cell line were evaluated for evidence of vascular mimicry. PAS staining, H&E staining, and CD31 immunohistochemistry were used to evaluate the presence and extent of mimicry as previously described (Folberg et al. 2000; Hendrix et al. 2003; Maniotis et al. 1999). Five random 40X fields from 2 different tumors (ten total fields) of each clone were scored for the number and size of areas with morphology consistent with mimicry. The criteria used was 1) PAS positive channels that contain red cells and fluid, 2) the absence of CD31 staining in these channels, and 3) the polarization of tumor cells on an indistinct or imperceptible matrix lining vascular channels with red cells and/or fluid and no evidence of endothelization or tumor cells lining vascular spaces with no evidence of a matrix.

Vascular leakage

To visualize vascular leakage in the primary tumor, 100 µl of Dextran Alexa 647, 10 kDa (1 mg/ml in PBS) (Life Technologies) were injected into mice by tail vein injection. Three minutes later, mice were perfused with 4% PFA. After fixation, tumors were harvested and put in 4% PFA overnight at 4 C. After this, samples were infiltrated with 20% sucrose overnight at 4 C. Tumors were frozen in OCT compound (Sakura Finetek) and 25 um thick sections were cut, washed, incubated with DAPI (1 mg/ml) (Sigma-Aldrich) and

mounted in ProLong Gold antifade reagent (Life Technologies). Sections were examined under the LSM 780 Confocal microscope (Zeiss).

An average of five to seven fields were taken from each sample. Each condition consisted of four to five animals (20 to 35 total fields). Images were quantified using ImageJ software (NIH). For quantifying fluorescence, the threshold of each picture was adjusted to the lowest possible value in the DAPI channel in order to measure total tissue area. The dextran threshold was fixed in each picture at a determined value based on the average intensity of all samples processed. The dextran-positive area was then normalized to the total tissue area in order to calculate the leakiness index.

qPCR for circulating tumor cells

Genomic DNA for blood was isolated using the QIAamp DNA Blood Mini Kit (Qiagen) and quantified using Prime Time qPCR assays (IDT). All samples were processed in triplicates. Each reaction consisted of 50 µl, containing 25 µl of iTaq Universal Supermix (BioRad), 2.5 µl of barcode primers and probe (Primer 1: 5'-ATC CAG AGG TTG ATT GTT CCA GAC GCG T-3', primer 2: 5'-CAG AAT CGT TGC CTG CAC ATC TTG GAA AC-3', FAM probe: 5'-/56-FAM/AAG GCT CGA /ZEN/GAC GTA GTC AGA CGT /3IABkFQ/-3'), 2.5 µl of housekeeping (NM_172901.2) primers and probe (Primer 1: 5'-GAC TTG TAA CGG GCA GGC AGA TTG TG-3', primer 2: 5'-GAG GTG TGG GTC ACC TCG ACA TC-3', HEX probe: 5'-/5HEX/CCG TGT CGC /ZEN/TCT GAA GGG CAA TAT /3IABkFQ/-3, IDT) and 20 µl of gDNA sample (100 ng). The cycling conditions were 1 cycle of denaturation at 95 C for 3 min, followed by 40 cycles of amplification (95 C for 15 sec, 68 C for 1 min). Quantitative PCR analysis was performed on the Eppendorf Mastercycler ep

realplex. All signals were quantified using the ΔCt method and were normalized to the levels of the housekeeping gene.

Tube formation assay

96-well plates were coated with 50 ul of growth factor reduced Matrigel (BD Biosciences) and 5,000 cells were re-suspended in EBM-2 media and plated in each well. All cells were plated in four replicates. Morphological studies were performed after 8 hours using the Zeiss Axio Observer inverted microscope.

Chapter contributions

Simon Knott, Mar Soto, Sara Gutiérrez-Ángel, Christina Hartl and Alissa Williams assisted with experiments and analysis. Chuck Harrell and Charles M. Perou helped with human expression data. John Wilkinson performed histological analysis.

Chapter 5: Conclusions and Future Perspectives

A transcriptional landscape of early human breast cancer

Very little is known about the genes that contribute to the development of DCIS and how they are altered to enable progression to invasive cancer. DCIS only progresses to invasive disease for less than half of the patients. Using laser capture microscopy and next-generation sequencing technology, we isolated DCIS and IDC lesions from ten patient biopsies. We utilized RNAseq methods that are specific for low input samples and generated transcriptional profiles. We then examined for clustered and unique expression changes across these patient samples in order to identify early breast cancer subtypes and differentially expressed genes between DCIS and IDC.

Using hierarchical clustering of all expressed genes in our dataset, we identified clusters of largely normal, DCIS and IDC samples. We were able to observe that samples from the same patient clustered together if they represent the same grade DCIS. However, there were a few small subgroups in the clustering that did not fit cleanly into groups with similar pathology. With more samples, we will be able to determine if these samples represent small but repeatable subgroups or outliers. Furthermore, we were able to identify distinct breast cancer subtypes among the DCIS and IDC samples and thus we may therefore be able to predict disease progression based on these subtypes. Lastly, we were able to identify candidate genes that are differentially expressed in DCIS vs normal ducts and DCIS vs IDC. These genes could be markers that predict DCIS or invasiveness. Profiling more samples, we will eventually be able to identify an invasion specific signature. It will be important to follow-up on these candidate genes using mouse models of DCIS, such as the DCIS cell lines Mcf10DCIS and Sum-225 (Behbod et al. 2009).

In the future, we are planning to extend this to hundreds of patient biopsies in order to provide a more detailed transcriptional landscape of DCIS and early IDC. The key factor that will allow us to determine changes that are crucial to malignant progression lies in the long-term profiles from patients that did and did not go on to develop invasive disease. This allows stratification of patients that do not need intervention as well as the identification of key targets that might prevent further progression. Currently, we are working with our collaborators at Duke University to receive all clinical annotations of our patient samples.

In addition, current literature suggests that the stromal compartment surrounding invasive breast cancer could play a role in the progression of the disease. The break from DCIS to invasive carcinoma could be a result, not of the cancerous cells themselves, but of the stromal compartment (any connective tissue cells such as fibroblasts or immune cells) shifting to accommodate the expansion of tumor cells. We are planning to profile these stromal compartments using RNAseq in order to see if we are able to pick up stromal signatures in these neighboring cell types and whether the stromal compartment is different between DCIS and IDC samples.

A transcriptional landscape of disseminated and circulating tumor cells

Metastasis is a stepwise mechanism starting with tumor cells invading through the basement membrane, following intravasation into the blood or lymphatic system, survival in the blood stream, extravasation into distant organ sites, and finally colonization. The RNAseq transcriptome data of the 4T1 breast cancer cell lines from primary tumor cells, CTCs and metastatic tumor cells in lung and liver represents a broad overview of the gene expression levels at different stages of cancer progression and metastasis. We were able to identify Ptpn22 as a novel metastatic inducer that is overexpressed in the CTC cells

compared to the primary tumor. Knock-down of Ptpn22 in 4T1 cells greatly reduced the metastatic burden to the lung when these cells are injected orthotopically into the mammary gland. In addition, we were able to detect a reduction in the number of CTCs in the blood upon Ptpn22 knock-down and thus believe that Ptpn22 is pro-survival for these cells. Future experiments will need to be conducted in order to validate the exact mode of action. Using protein arrays, we are planning to identify substrates for Ptpn22. Another alternative to identify substrates of Ptpn22 might be substrate-trapping, in which catalytic inactive Ptpn22 mutants are designed, which will bind to their tyrosine phosphorylated substrates (Flint et al. 1997). Most commonly the active cysteine residue of the signature motif is mutated to a serine residue. Substrates can be identified *in vivo* through co-immunoprecipitation or *in vitro* through GST pulldown and mass spectrometry. Moreover, overexpression of Ptpn22 in weakly metastatic cells such as 4T07 cells should be conducted in order to see if there is an increase in metastatic burden. In order to validate that overexpression of Ptpn22 in the CTC compartment is pro-survival *in vivo*, we are planning to overexpress pro-survival Bcl-2 in Pptn22 knocked-down 4T1 cells in order to see if this rescues the reduction of metastatic burden in the lung.

The identification and characterization of differentially expressed genes in different stages of cancer progression, such as Ptpn22, will help identify novel drug target that inhibit metastasis formation. As low input RNAseq protocols are steadily improving, it is now feasible to do single-cell RNAseq libraries. This will further aid in the identification of drug targets, since we will be able to assess the heterogeneity of the CTC population.

A murine model of tumor heterogeneity

The vast majority of recent studies focusing on tumor heterogeneity and how it influences metastatic disease progression have been retrospective in nature. Although, these studies often list putative drivers of specific clonal phenotypes, they are not able to validate them through functional genomics. We have developed a murine model of tumor heterogeneity that can be perturbed genetically to validate drivers of specific clonal behaviors. This model has the potential to elucidate drivers of all stages of tumor formation and metastasis.

With this new technology we have found that a small subset of clones within the primary tumor are able to enter the cardiovascular system. These clones are not abundant in the primary tumor, and their ability to infiltrate the vascular system is not indicative of their ability to enter the lymphatic system. Furthermore, we demonstrated that these cells share the unique ability to perform vascular mimicry (VM), both *in vivo* and *in vitro*. VM is a phenomenon wherein highly aggressive tumor cells differentiate to have endothelial-like phenotypes, allowing them to form channels that emanate from the tumor to the vasculature to distribute blood to hypoxic regions of the tumor.

By performing expression profiling on these specialized cells, we have identified two serine protease inhibitors (Serpine2 and Slpi) that are both necessary and sufficient for the capacity of tumor cells to perform VM. We showed that these genes confer the same phenotypes in basal and claudin-low human breast cancer cell lines. Finally, through the analysis of clinical data, we showed that these genes are significantly up-regulated in breast cancer patients with relapse to the lung, as compared to non-relapsed patients. This represents a novel technology for identifying clinically relevant drivers of each stage of metastatic disease and highlights viable targets for the prevention of tumor cell

dissemination. Perhaps most importantly, these findings provide a logical basis for the selection of metastatic potential. Those changes in gene expression that feed the primary tumor with blood via vascular mimicry produce characteristics that promote metastasis to secondary sites.

Vascular mimicry was initially identified in aggressive melanoma cells that were able to form tubular networks in three-dimensional culture systems (Maniotis et al. 1999). These observations were extended to melanoma xenograft models and biopsies of human aggressive melanoma, and then to other tumor types such as breast cancer (Shirakawa et al. 2001). Later, investigators were able to show that these tubular structures contained viable blood flow using a xenograft model of inflammatory breast cancer (Shirakawa et al. 2002). A recent study conducted a meta-analysis of vascular mimicry and its association with poor prognosis in human cancer patients (Cao et al. 2013). Interestingly, vascular mimicry cancer patients show a poor 5-year overall survival (31%) compared to vascular mimicry negative patients (56%). Moreover metastatic melanoma patients had higher vascular mimicry rates (45%) than cancer patients with primary melanoma (23%). Thus it seems that vascular mimicry could contribute to metastasis formation in human cancer patients. Because the survival benefits of anti-angiogenic drugs have been modest so far in the clinic, it will be important to consider vascular mimicry in the design of anti-vascular therapies (Garber 2002). For this, we are envisioning to combine anti-angiogenic drugs with knock-down of Serpine2 or Slpi in mice in order to see if it further reduces the lung metastatic burden.

Slpi and Serpine2 have anti-coagulation effects, since these secreted proteins can bind and inhibit thrombin. We hypothesize that the vascular mimicry network has more leakiness than endothelial-lined vasculature and that the direct consequence is a higher rate of blood coagulation in these tubular structures. However, because of high expression of

Slpi and Serpine2, these cells are able to inhibit coagulation and can access free flowing blood within the vascular mimicry network. This fuels the primary tumor with oxygen and nutrients, and allows these cells to be drawn into the blood stream as circulating tumor cells, which can lead to metastasis formation. A recent publication from the Massagué group, identified other Serpins to be promoting cancer cell survival and angiogenesis once they have reached the brain during metastasis formation (Valiente et al. 2014). Interestingly, they also identified Serpine2 as one of the genes up-regulated in brain metastases compared to the primary tumor. However, they validated these genes through knock-down studies using intravenous injections into the blood stream in order to assess brain metastases. In these experiments, Serpine2 did not validate for the authors, since there was requirement for intravasation of these cells. Thus, it seems that Serpins in general have important functions in regulating metastasis formation and that approaches should be identified that will allow the targeting of multiple Serpins at once.

Tumor heterogeneity is a major challenge in current clinical cancer therapy. It is a direct consequence of the adjustable nature of tumor cells, which allows them to adapt to a range of different environmental influences. Over the course of cancer progression, this leads to the development of a complex hierarchical structure of tumor subpopulations with various characteristics, such as in this case subpopulations that are efficient in VM. Current therapy often cannot take on the vast number of diverse cellular clones, which ultimately result in the dominance of a few clones. In order to improve treatment options for cancer patients, it is necessary to establish the ability to quickly characterize the impact of genomic alterations on the tumor and metastasis. We anticipate that our method is widely applicable to a variety of cancer types and hope that it can eventually be applied online to patient derived material. One future approach might be to take a small biopsy from a patient tumor,

introduce barcodes and then single-sort individual clonal lines. These clonal lines are then mixed and injected into mice in order to identify highly metastatic or resistant cell lineages. After elucidation of these vulnerabilities from these patient derived materials, one could adjust and personalize treatment strategies. In addition, this technology simultaneously allows the identification of novel drug targets that are specific to certain cell lineages.

Another important aspect of heterogeneity is cooperativity between different clones. This is nicely shown in a study that uses Wnt1 induced mammary tumors in mice, which are composed of mixtures of basal and luminal cancer cells (Cleary et al. 2014). Interestingly, not luminal nor basal cancer cells are able to form mammary tumors by themselves, only in combination of both populations, they are able to induce tumor formation at a high rate. The authors were able to show that this is due to paracrine interaction of Wnt1 between both individual populations. In another study, which also highlights clonal cooperativity, the authors used the human cancer cell line MDA-MB-468 to generate a panel of 18 sub-clones that differed in overexpression of single secreted factors (Marusyk et al. 2014). They used these clones to generate monoclonal or polyclonal tumors in order to conduct clonal competition assays *in vivo*. They were able to show that tumor growth can be driven by a minor clonal population that expresses IL-11, which enhanced the proliferation of all tumor cells. Thus, in regards to our own heterogeneity model, it would be interesting to carry out competition assays with different combinations of clones. Clone T might for example be only effective in intravasating into the blood in the presence of a second clone.

Identification of extravasation specific genes

In our murine model of tumor heterogeneity we were able to identify two clones, clone E and T, which were able to disseminate with equal efficiency to the cardiovascular

system. However, the majority of metastatic cells in the lung were composed solely of clone T. Thus, clone T is far more efficient in the extravasation process from the cardiovascular system to the lung, compared to clone E. We can exclude that this is a mere colonization phenotype of clone T since the number of secondary barcodes in clone T far exceeds that of clone E. We would like to exploit this observations and use clone T in an shRNA screen setting to identify genes that are necessary for extravasation. For this, we want to compare the RNAseq profiles of clone E and T and identify genes that are overexpressed in clone T compared to clone E. These genes will be used in an shRNA screen in clone T and we will score the ability of these cells to extravasate from the blood to the lung, when injected into the tail vein of mice. We anticipate that this will result in the identification of novel genes that are specific for the extravasation process. By co-targeting one of these extravasation specific genes with Serpine2/Slpi, we think that we can drastically reduce the lung metastastic burden even more since we will be targeting both the intravasation and extravasation step of metastasis.

References

Abbracchio MP, Boeynaems J-M, Barnard EA, Boyer JL, Kennedy C, Miras-Portugal MT, King BF, Gachet C, Jacobson KA, Weisman GA, et al. 2003. Characterization of the UDP-glucose receptor (re-named here the P2Y14 receptor) adds diversity to the P2Y receptor family. *Trends Pharmacol Sci* **24**: 52–55.

Aguirre-Ghiso JA. 2007. Models, mechanisms and clinical evidence for cancer dormancy. *Nat Rev Cancer* **7**: 834–846.

Alix-Panabières C, Pantel K. 2014. Technologies for detection of circulating tumor cells: facts and vision. *Lab Chip* **14**: 57–62.

Allred DC. 2010. Ductal carcinoma in situ: terminology, classification, and natural history. *J Natl Cancer Inst Monographs* **2010**: 134–138.

Aslakson CJ, Miller FR. 1992. Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. *Cancer Res* **52**: 1399–1405.

Auer PL, Doerge RW. 2010. Statistical design and analysis of RNA sequencing data. *Genetics* **185**: 405–416.

Aychek T, Miller K, Sagi-Assif O, Levy-Nissenbaum O, Israeli-Amit M, Pasmanik-Chor M, Jacob-Hirsch J, Amariglio N, Rechavi G, Witz IP. 2008. E-selectin regulates gene expression in metastatic colorectal carcinoma cells and enhances HMGB1 release. *Int J Cancer* **123**: 1741–1750.

Bach LA, Bentzen SM, Alsner J, Christiansen FB. 2001. An evolutionary-game model of tumour-cell interactions: possible relevance to gene therapy. *Eur J Cancer* **37**: 2116–2120.

Bandyopadhyay S, Zhan R, Chaudhuri A, Watabe M, Pai SK, Hirota S, Hosobe S, Tsukada T, Miura K, Takano Y, et al. 2006. Interaction of KAI1 on tumor cells with DARC on vascular endothelium leads to metastasis suppression. *Nat Med* **12**: 933–938.

Barnes NLP, Ooi JL, Yarnold JR, Bundred NJ. 2012. Ductal carcinoma in situ of the breast. *BMJ* **344**: e797.

Beck B, Blanpain C. 2013. Unravelling cancer stem cell potential. *Nat Rev Cancer* **13**: 727–738.

Behbod F, Kittrell FS, LaMarca H, Edwards D, Kerbawy S, Heestand JC, Young E, Mukhopadhyay P, Yeh H-W, Allred DC, et al. 2009. An intraductal human-in-mouse transplantation model mimics the subtypes of ductal carcinoma in situ. *Breast Cancer Res* **11**: R66.

Benbrook DM, Masamha CP. 2011. The pro-survival function of Akt kinase can be overridden or altered to contribute to induction of apoptosis. *Curr Cancer Drug Targets* **11**: 586–599.

Bidwell BN, Slaney CY, Withana NP, Forster S, Cao Y, Loi S, Andrews D, Mikeska T, Mangan NE, Samarajiwa SA, et al. 2012. Silencing of Irf7 pathways in breast cancer cells promotes bone metastasis through immune escape. *Nat Med* **18**: 1224–1231.

Bissell MJ, Labarge MA. 2005. Context, tissue plasticity, and cancer: are tumor stem cells also regulated by the microenvironment? *Cancer Cell* **7**: 17–23.

Bissell MJ, Radisky D. 2001. Putting tumours in context. *Nat Rev Cancer* **1**: 46–54.

Bobek V, Kovarík J. 2004. Antitumor and antimetastatic effect of warfarin and heparins. *Biomed Pharmacother* **58**: 213–219.

Bos PD, Zhang XH-F, Nadal C, Shu W, Gomis RR, Nguyen DX, Minn AJ, van de Vijver MJ, Gerald WL, Foekens JA, et al. 2009. Genes that mediate breast cancer metastasis to the brain. *Nature* **459**: 1005–1009.

Bouton M-C, Boulaftali Y, Richard B, Arocas V, Michel J-B, Jandrot-Perrus M. 2012. Emerging role of serpinE2/protease nexin-1 in hemostasis and vascular biology. *Blood* **119**: 2452–2457.

Braun S, Kentenich C, Janni W, Hepp F, de Waal J, Willgeroth F, Sommer H, Pantel K. 2000. Lack of effect of adjuvant chemotherapy on the elimination of single dormant tumor cells in bone marrow of high-risk breast cancer patients. *J Clin Oncol* **18**: 80–86.

Broeks A, Braaf LM, Huseinovic A, Nooijen A, Urbanus J, Hogervorst FBL, Schmidt MK, Klijn JGM, Russell NS, Van Leeuwen FE, et al. 2007. Identification of women with an increased risk of developing radiation-induced breast cancer: a case only study. *Breast Cancer Res* **9**: R26.

Brown TM, Fee E. 2006. *Rudolf Carl Virchow: medical scientist, social reformer, role model.*

Brownlie RJ, Miosge LA, Vassilakos D, Svensson LM, Cope A, Zamoyska R. 2012. Lack of the phosphatase PTPN22 increases adhesion of murine regulatory T cells to improve their immunosuppressive function. *Sci Signal* **5**: ra87–ra87.

Buchholz M, Biebl A, Neesse A, Wagner M, Iwamura T, Leder G, Adler G, Gress TM. 2003. SERPINE2 (protease nexin I) promotes extracellular matrix production and local invasion of pancreatic tumors in vivo. *Cancer Res* **63**: 4945–4951.

Camerer E, Qazi AA, Duong DN, Cornelissen I, Advincula R, Coughlin SR. 2004. Platelets, protease-activated receptors, and fibrinogen in hematogenous metastasis. *Blood* **104**: 397–401.

Cancer Genome Atlas Network. 2012. Comprehensive molecular portraits of human breast tumours. *Nature* **490**: 61–70.

Cann GM, Gulzar ZG, Cooper S, Li R, Luo S, Tat M, Stuart S, Schroth G, Srinivas S, Ronaghi M, et al. 2012. mRNA-Seq of single prostate cancer circulating tumor cells reveals recapitulation of gene expression and pathways found in prostate cancer. *PLoS ONE* **7**: e49144.

Cao Z, Bao M, Miele L, Sarkar FH, Wang Z, Zhou Q. 2013. Tumour vasculogenic mimicry is associated with poor prognosis of human cancer patients: a systemic review and meta-analysis. *Eur J Cancer* **49**: 3914–3923.

Casares N, Arribillaga L, Sarobe P, Dotor J, Lopez-Diaz de Cerio A, Melero I, Prieto J, Borrás-Cuesta F, Lasarte JJ. 2003. CD4+/CD25+ regulatory cells inhibit activation of tumor-primed CD4+ T cells with IFN-gamma-dependent antiangiogenic activity, as well as long-lasting tumor immunity elicited by peptide vaccination. *J Immunol* **171**: 5931–5939.

Cerutti JM, Oler G, Michaluart P, Delcelo R, Beaty RM, Shoemaker J, Riggins GJ. 2007. Molecular profiling of matched samples identifies biomarkers of papillary thyroid carcinoma lymph node metastasis. *Cancer Res* **67**: 7885–7892.

Chan SW, Lim CJ, Guo K, Ng CP, Lee I, Hunziker W, Zeng Q, Hong W. 2008. A role for TAZ in migration, invasion, and tumorigenesis of breast cancer cells. *Cancer Res* **68**: 2592–2598.

Cifuentes N, Pickren JW. 1979. Metastases from carcinoma of mammary gland: an autopsy study. *J Surg Oncol* **11**: 193–205.

Cleary AS, Leonard TL, Gestl SA, Gunther EJ. 2014. Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. *Nature* **508**: 113–117.

Cohen S, Dadi H, Shaoul E, Sharfe N, Roifman CM. 1999. Cloning and characterization of a lymphoid-specific, inducible human protein tyrosine phosphatase, Lyp. *Blood* **93**: 2013–2024.

Coomber BL, Yu JL, Fathers KE, Plumb C, Rak JW. 2003. Angiogenesis and the role of epigenetics in metastasis. *Clin Exp Metastasis* **20**: 215–227.

Cordier AC, Haumont SM. 1980. Development of thymus, parathyroids, and ultimobranchial bodies in NMRI and nude mice. *Am J Anat* **157**: 227–263.

Corver WE, Middeldorp A, Haar ter NT, Jordanova ES, van Puijenbroek M, van Eijk R, Cornelisse CJ, Fleuren GJ, Morreau H, Oosting J, et al. 2008. Genome-wide allelic state analysis on flow-sorted tumor fractions provides an accurate measure of chromosomal aberrations. *Cancer Res* **68**: 10333–10340.

Coussens LM, Raymond WW, Bergers G, Laig-Webster M, Behrendtsen O, Werb Z, Caughey GH, Hanahan D. 1999. Inflammatory mast cells up-regulate angiogenesis during squamous epithelial carcinogenesis. *Genes Dev* **13**: 1382–1397.

Curiel TJ, Coukos G, Zou L, Alvarez X, Cheng P, Mottram P, Evdemon-Hogan M, Conejo-Garcia JR, Zhang L, Burow M, et al. 2004. Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. *Nat Med* **10**: 942–949.

DeNardo DG, Barreto JB, Andreu P, Vasquez L, Tawfik D, Kolhatkar N, Coussens LM. 2009. CD4(+) T cells regulate pulmonary metastasis of mammary carcinomas by enhancing

protumor properties of macrophages. *Cancer Cell* **16**: 91–102.

DeSantis C, Ma J, Bryan L, Jemal A. 2014. Breast cancer statistics, 2013. *CA Cancer J Clin* **64**: 52–62.

Dexter DL, Kowalski HM, Blazar BA, Fligiel Z, Vogel R, Heppner GH. 1978. Heterogeneity of tumor cells from a single mouse mammary tumor. *Cancer Res* **38**: 3174–3181.

Dialynas DP, Wilde DB, Marrack P, Pierres A, Wall KA, Havran W, Otten G, Loken MR, Pierres M, Kappler J. 1983. Characterization of the murine antigenic determinant, designated L3T4a, recognized by monoclonal antibody GK1.5: expression of L3T4a by functional T cell clones appears to correlate primarily with class II MHC antigen-reactivity. *Immunol Rev* **74**: 29–56.

Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, et al. 2010. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**: 999–1005.

Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, McLellan MD, et al. 2012. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**: 506–510.

Donker M, Litière S, Werutsky G, Julien J-P, Fentiman IS, Agresti R, Rouanet P, de Lara CT, Bartelink H, Duez N, et al. 2013. Breast-conserving treatment with or without radiotherapy in ductal carcinoma In Situ: 15-year recurrence rates and outcome after a recurrence, from the EORTC 10853 randomized phase III trial. *J Clin Oncol* **31**: 4054–4059.

Egeblad M, Nakasone ES, Werb Z. 2010. Tumors as organs: complex tissues that interface with the entire organism. *Dev Cell* **18**: 884–901.

Emadi Baygi M, Soheili ZS, Schmitz I, Sameie S, Schulz WA. 2010. Snail regulates cell survival and inhibits cellular senescence in human metastatic prostate cancer cell lines. *Cell Biol Toxicol* **26**: 553–567.

Emberley ED, Alowami S, Snell L, Murphy LC, Watson PH. 2004. S100A7 (psoriasin) expression is associated with aggressive features and alteration of Jab1 in ductal carcinoma in situ of the breast. *Breast Cancer Res* **6**: R308–15.

Esserman LJ, Berry DA, Cheang MCU, Yau C, Perou CM, Carey L, DeMichele A, Gray JW, Conway-Dorsey K, Lenburg ME, et al. 2012. Chemotherapy response and recurrence-free survival in neoadjuvant breast cancer depends on biomarker profiles: results from the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657). *Breast Cancer Res Treat* **132**: 1049–1062.

Fakhrejahani E, Toi M. 2012. Tumor angiogenesis: pericytes and maturation are not to be ignored. *J Oncol* **2012**: 261750–10.

Fein MR, Egeblad M. 2013. Caught in the act: revealing the metastatic process by live imaging. *Dis Model Mech* **6**: 580–593.

Fialkow PJ. 1979. Clonal origin of human tumors. *Annu Rev Med* **30**: 135–143.

Fidler IJ, Kripke ML. 1977. Metastasis results from preexisting variant cells within a malignant tumor. *Science* **197**: 893–895.

Fisher R, Pusztai L, Swanton C. 2013. Cancer heterogeneity: implications for targeted therapeutics. *Br J Cancer* **108**: 479–485.

Flint AJ, Tiganis T, Barford D, Tonks NK. 1997. Development of "substrate-trapping" mutants to identify physiological substrates of protein tyrosine phosphatases. *Proc Natl Acad Sci USA* **94**: 1680–1685.

Folberg R, Hendrix MJ, Maniotis AJ. 2000. Vasculogenic mimicry and tumor angiogenesis. *Am J Pathol* **156**: 361–381.

Francescone RA, Faibish M, Shao R. 2011. A Matrigel-based tube formation assay to assess the vasculogenic activity of tumor cells. *J Vis Exp* e3040–e3040.

Fridlender ZG, Sun J, Kim S, Kapoor V, Cheng G, Ling L, Worthen GS, Albelda SM. 2009. Polarization of tumor-associated neutrophil phenotype by TGF-beta: "N1" versus "N2" TAN. *Cancer Cell* **16**: 183–194.

Garber K. 2002. Angiogenesis inhibitors suffer new setback. *Nature biotechnology*, November.

Gerlinger M, Swanton C. 2010. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *Br J Cancer* **103**: 1139–1143.

Giampieri S, Manning C, Hooper S, Jones L, Hill CS, Sahai E. 2009. Localized and reversible TGFbeta signalling switches breast cancer cells from cohesive to single cell motility. *Nat Cell Biol* **11**: 1287–1296.

Girardi M, Oppenheim DE, Steele CR, Lewis JM, Glusac E, Filler R, Hobby P, Sutton B, Tigelaar RE, Hayday AC. 2001. Regulation of cutaneous malignancy by gammadelta T cells. *Science* **294**: 605–609.

Gonzalez-Angulo AM, Iwamoto T, Liu S, Chen H, Do K-A, Hortobagyi GN, Mills GB, Meric-Bernstam F, Symmans WF, Pusztai L. 2012. Gene expression, molecular class changes, and pathway analysis after neoadjuvant systemic therapy for breast cancer. *Clin Cancer Res* **18**: 1109–1119.

Greaves M, Maley CC. 2012. Clonal evolution in cancer. *Nature* **481**: 306–313.

Greenburg G, Hay ED. 1982. Epithelia suspended in collagen gels can lose polarity and express characteristics of migrating mesenchymal cells. *J Cell Biol* **95**: 333–339.

Grivennikov S, Karin E, Terzic J, Mucida D, Yu G-Y, Vallabhapurapu S, Scheller J, Rose-John S, Cheroutre H, Eckmann L, et al. 2009. IL-6 and Stat3 are required for survival of intestinal epithelial cells and development of colitis-associated cancer. *Cancer Cell* **15**: 103–113.

Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, Lander ES. 2011. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**: 633–644.

Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. *Cell* **144**: 646–674.

Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. *Cell* **100**: 57–70.

Harrell JC, Prat A, Parker JS, Fan C, He X, Carey L, Anders C, Ewend M, Perou CM. 2012. Genomic analysis identifies unique signatures predictive of brain, lung, and liver relapse. *Breast Cancer Res Treat* **132**: 523–535.

Hendrix MJC, Seftor EA, Hess AR, Seftor REB. 2003. Vasculogenic mimicry and tumour-cell plasticity: lessons from melanoma. *Nat Rev Cancer* **3**: 411–421.

Hibi T, Mori T, Fukuma M, Yamazaki K, Hashiguchi A, Yamada T, Tanabe M, Aiura K, Kawakami T, Ogiwara A, et al. 2009. Synuclein-gamma is closely involved in perineural invasion and distant metastasis in mouse models and is a novel prognostic factor in pancreatic cancer. *Clin Cancer Res* **15**: 2864–2871.

Hotz B, Arndt M, Dullat S, Bhargava S, Buhr H-J, Hotz HG. 2007. Epithelial to mesenchymal transition: expression of the regulators snail, slug, and twist in pancreatic cancer. *Clin Cancer Res* **13**: 4769–4776.

Hovey RM, Chu L, Balazs M, DeVries S, Moore D, Sauter G, Carroll PR, Waldman FM. 1998. Genetic alterations in primary bladder cancers and their metastases. *Cancer Res* **58**: 3555–3560.

Iiizumi M, Bandyopadhyay S, Watabe K. 2007. Interaction of Duffy antigen receptor for chemokines and KAI1: a critical step in metastasis suppression. *Cancer Res* **67**: 1411–1414.

Inoue S, Leitner WW, Golding B, Scott D. 2006. Inhibitory effects of B cells on antitumor immunity. *Cancer Res* **66**: 7741–7747.

Israeli O, Gotlieb WH, Friedman E, Korach J, Friedman E, Goldman B, Zeltzer A, Zeltser A, Ben-Baruch G, Rienstein S, et al. 2004. Genomic analyses of primary and metastatic serous epithelial ovarian cancer. *Cancer Genet Cytogenet* **154**: 16–21.

Jia T, Liu YE, Liu J, Shi YE. 1999. Stimulation of breast cancer invasion and metastasis by synuclein gamma. *Cancer Res* **59**: 742–747.

Kaplan RN, Riba RD, Zacharoulis S, Bramley AH, Vincent L, Costa C, MacDonald DD, Jin DK, Shido K, Kerns SA, et al. 2005. VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature* **438**: 820–827.

Karrison TG, Ferguson DJ, Meier P. 1999. Dormancy of mammary carcinoma after mastectomy. *J Natl Cancer Inst* **91**: 80–85.

Keats JJ, Chesi M, Egan JB, Garbitt VM, Palmer SE, Braggio E, Van Wier S, Blackburn PR,

Baker AS, Dispenzieri A, et al. 2012. Clonal competition with alternating dominance in multiple myeloma. *Blood* **120**: 1067–1076.

Knight SC, Hunt R, Dore C, Medawar PB. 1985. Influence of dendritic cells on tumor growth. *Proc Natl Acad Sci USA* **82**: 4495–4497.

Komaki K, Sano N, Tangoku A. 2006. Problems in histological grading of malignancy and its clinical significance in patients with operable breast cancer. *Breast Cancer* **13**: 249–253.

Krebs MG, Metcalf RL, Carter L, Brady G, Blackhall FH, Dive C. 2014. Molecular analysis of circulating tumour cells-biology and biomarkers. *Nat Rev Clin Oncol* **11**: 129–144.

Kreso A, O'Brien CA, van Galen P, Gan OI, Notta F, Brown AMK, Ng K, Ma J, Wienholds E, Dunant C, et al. 2013. Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science* **339**: 543–548.

Krishnamurthy S, Bischoff F, Ann Mayer J, Wong K, Pham T, Kuerer H, Lodhi A, Bhattacharyya A, Hall C, Lucci A. 2013. Discordance in HER2 gene amplification in circulating and disseminated tumor cells in patients with operable breast cancer. *Cancer Med* **2**: 226–233.

Ledbetter JA, Herzenberg LA. 1979. Xenogeneic monoclonal antibodies to mouse lymphoid differentiation antigens. *Immunol Rev* **47**: 63–90.

Liu W, Laitinen S, Khan S, Vihinen M, Kowalski J, Yu G, Chen L, Ewing CM, Eisenberger MA, Carducci MA, et al. 2009. Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. *Nat Med* **15**: 559–565.

Lohr JG, Stojanov P, Carter SL, Cruz-Gordillo P, Lawrence MS, Auclair D, Sougnez C, Knoechel B, Gould J, Saksena G, et al. 2014. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell* **25**: 91–101.

Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J, et al. 2003. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res* **13**: 2291–2305.

Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, et al. 2004. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350**: 2129–2139.

MAKINO S. 1956. Further evidence favoring the concept of the stem cell in ascites tumors of rats. *Ann N Y Acad Sci* **63**: 818–830.

Mani SA, Guo W, Liao M-J, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, et al. 2008. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**: 704–715.

Maniotis AJ, Folberg R, Hess A, Seftor EA, Gardner LM, Pe'er J, Trent JM, Meltzer PS, Hendrix MJ. 1999. Vascular channel formation by human melanoma cells in vivo and in

vitro: vasculogenic mimicry. *Am J Pathol* **155**: 739–752.

Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD, et al. 2009. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* **361**: 1058–1066.

Marusyk A, Tabassum DP, Altrock PM, Almendro V, Michor F, Polyak K. 2014. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature*.

Masuda K, Kamimura T, Watanabe K, Suga T, Kanesaki M, Takeuchi A, Imaizumi A, Suzuki Y. 1995. Pharmacological activity of the C-terminal and N-terminal domains of secretory leukoprotease inhibitor in vitro. *Br J Pharmacol* **115**: 883–888.

Matsubayashi J, Takanashi M, Oikawa K, Fujita K, Tanaka M, Xu M, De Blasi A, Bouvier M, Kinoshita M, Kuroda M, et al. 2008. Expression of G protein-coupled receptor kinase 4 is associated with breast cancer tumourigenesis. *J Pathol* **216**: 317–327.

Mayordomo JI, Zorina T, Storkus WJ, Zitvogel L, Celluzzi C, Falo LD, Melief CJ, Ildstad ST, Kast WM, Deleo AB. 1995. Bone marrow-derived dendritic cells pulsed with synthetic tumour peptides elicit protective and therapeutic antitumour immunity. *Nat Med* **1**: 1297–1302.

McCulloch P, George WD. 1987. Warfarin inhibition of metastasis: the role of anticoagulation. *Br J Surg* **74**: 879–883.

Meacham CE, Morrison SJ. 2013. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**: 328–337.

Meyerson AF, Lessing JN, Itakura K, Hylton NM, Wolverton DE, Joe BN, Esserman LJ, Hwang ES. 2011. Outcome of long term active surveillance for estrogen receptor-positive ductal carcinoma in situ. *Breast* **20**: 529–533.

Miller BE, McInerney D, Jackson D, Miller FR. 1986. Metabolic cooperation between mouse mammary tumor subpopulations in three-dimensional collagen gel cultures. *Cancer Res* **46**: 89–93.

Miller BE, Miller FR, Wilburn DJ, Heppner GH. 1987. Analysis of tumour cell composition in tumours composed of paired mixtures of mammary tumour cell lines. *Br J Cancer* **56**: 561–569.

Miller BE, Roi LD, Howard LM, Miller FR. 1983a. Quantitative selectivity of contact-mediated intercellular communication in a metastatic mouse mammary tumor line. *Cancer Res* **43**: 4102–4107.

Miller FR, Miller BE, Heppner GH. 1983b. Characterization of metastatic heterogeneity among subpopulations of a single mouse mammary tumor: heterogeneity in phenotypic stability. *Invasion Metastasis* **3**: 22–31.

Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, Viale A, Olshen AB, Gerald WL, Massagué J. 2005. Genes that mediate breast cancer metastasis to lung. *Nature* **436**: 518–

524.

Mousa SA. 2006. Role of current and emerging antithrombotics in thrombosis and cancer. *Drugs Today* **42**: 331–350.

Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, Downing JR. 2008. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* **322**: 1377–1380.

Müller A, Homey B, Soto H, Ge N, Catron D, Buchanan ME, McClanahan T, Murphy E, Yuan W, Wagner SN, et al. 2001. Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410**: 50–56.

Müller V, Riethdorf S, Rack B, Janni W, Fasching PA, Solomayer E, Aktas B, Kasimir-Bauer S, Pantel K, Fehm T, et al. 2012. Prognostic impact of circulating tumor cells assessed with the CellSearch System™ and AdnaTest Breast™ in metastatic breast cancer patients: the DETECT study. *Breast Cancer Res* **14**: R118.

Nakata B, Fukunaga S, Noda E, Amano R, Yamada N, Hirakawa K. 2008. Chemokine receptor CCR7 expression correlates with lymph node metastasis in pancreatic cancer. *Oncology* **74**: 69–75.

Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, et al. 2011. Tumour evolution inferred by single-cell sequencing. *Nature* **472**: 90–94.

Navin N, Krasnitz A, Rodgers L, Cook K, Meth J, Kendall J, Riggs M, Eberling Y, Troge J, Grubor V, et al. 2010. Inferring tumor progression from genomic heterogeneity. *Genome Res* **20**: 68–80.

Neiva KG, Warner KA, Campos MS, Zhang Z, Moren J, Danciu TE, Nör JE. 2014. Endothelial cell-derived interleukin-6 regulates tumor growth. *BMC Cancer* **14**: 99.

Newell EW, Davis MM. 2014. Beyond model antigens: high-dimensional methods for the analysis of antigen-specific T cells. *Nat Biotechnol* **32**: 149–157.

Nguyen DX, Bos PD, Massagué J. 2009. Metastasis: from dissemination to organ-specific colonization. *Nat Rev Cancer* **9**: 274–284.

Nielsen BS, Rank F, Illemann M, Lund LR, Danø K. 2007. Stromal cells associated with early invasive foci in human mammary ductal carcinoma in situ coexpress urokinase and urokinase receptor. *Int J Cancer* **120**: 2086–2095.

Nieswandt B, Hafner M, Echtenacher B, Männel DN. 1999. Lysis of tumor cells by natural killer cells in mice is impeded by platelets. *Cancer Res* **59**: 1295–1300.

Nolan-Stevaux O, Tedesco D, Ragan S, Makhanov M, Chenchik A, Ruefli-Brasse A, Quon K, Kassner PD. 2013. Measurement of Cancer Cell Growth Heterogeneity through Lentiviral Barcoding Identifies Clonal Dominance as a Characteristic of In Vivo Tumor Engraftment. *PLoS ONE* **8**: e67316.

Nowell PC. 1976. The clonal evolution of tumor cell populations. *Science* **194**: 23–28.

Olive KP, Jacobetz MA, Davidson CJ, Gopinathan A, McIntyre D, Honess D, Madhu B, Goldgraben MA, Caldwell ME, Allard D, et al. 2009. Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science* **324**: 1457–1461.

Orimo A, Gupta PB, Sgroi DC, Arenzana-Seisdedos F, Delaunay T, Naeem R, Carey VJ, Richardson AL, Weinberg RA. 2005. Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* **121**: 335–348.

Ostrand-Rosenberg S, Pulaski BA, Armstrong TD, Clements VK. 1998. Immunotherapy of established tumor with MHC class II and B7.1 cell-based tumor vaccines. *Adv Exp Med Biol* **451**: 259–264.

Otterbach F, Callies R, Adamzik M, Kimmig R, Siffert W, Schmid KW, Bankfalvi A. 2010. Aquaporin 1 (AQP1) expression is a novel characteristic feature of a particularly aggressive subgroup of basal-like breast carcinomas. *Breast Cancer Res Treat* **120**: 67–76.

Pantel K, Brakenhoff RH. 2004. Dissecting the metastatic cascade. *Nat Rev Cancer* **4**: 448–456.

Pantel K, Schlimok G, Braun S, Kutter D, Lindemann F, Schaller G, Funke I, Izbicki JR, Riethmüller G. 1993. Differential expression of proliferation-associated molecules in individual micrometastatic carcinoma cells. *J Natl Cancer Inst* **85**: 1419–1424.

Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, et al. 2009. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* **27**: 1160–1167.

Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al. 2000. Molecular portraits of human breast tumours. *Nature* **406**: 747–752.

Petersson S, Bylander A, Yhr M, Enerbäck C. 2007. S100A7 (Psoriasin), highly expressed in ductal carcinoma in situ (DCIS), is regulated by IFN-gamma in mammary epithelial cells. *BMC Cancer* **7**: 205.

Pfitzenmaier J, Ellis WJ, Arfman EW, Hawley S, McLaughlin PO, Lange PH, Vessella RL. 2006. Telomerase activity in disseminated prostate cancer cells. *BJU Int* **97**: 1309–1313.

Porter D, Lahti-Domenici J, Keshaviah A, Bae YK, Argani P, Marks J, Richardson A, Cooper A, Strausberg R, Riggins GJ, et al. 2003. Molecular markers in ductal carcinoma in situ of the breast. *Mol Cancer Res* **1**: 362–375.

Pulaski BA, Ostrand-Rosenberg S. 1998. Reduction of established spontaneous mammary carcinoma metastases following immunotherapy with major histocompatibility complex class II and B7.1 cell-based tumor vaccines. *Cancer Res* **58**: 1486–1493.

Pulaski BA, Terman DS, Khan S, Muller E, Ostrand-Rosenberg S. 2000. Cooperativity of

Staphylococcal aureus enterotoxin B superantigen, major histocompatibility complex class II, and CD80 for immunotherapy of advanced spontaneous metastases in a clinically relevant postoperative mouse breast cancer model. *Cancer Res* **60**: 2710–2715.

Ramsköld D, Luo S, Wang Y-C, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF, Laurent LC, et al. 2012. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* **30**: 777–782.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Saunders NA, Simpson F, Thompson EW, Hill MM, Endo-Munoz L, Leggatt G, Minchin RF, Guminski A. 2012. Role of intratumoural heterogeneity in cancer drug resistance: molecular and clinical perspectives. *EMBO Mol Med* **4**: 675–684.

Schuh A, Becq J, Humphray S, Alexa A, Burns A, Clifford R, Feller SM, Grocock R, Henderson S, Khrebtukova I, et al. 2012. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* **120**: 4191–4196.

Shackleton M, Quintana E, Fearon ER, Morrison SJ. 2009. Heterogeneity in cancer: cancer stem cells versus clonal evolution. *Cell* **138**: 822–829.

Shah NP, Nicoll JM, Nagar B, Gorre ME, Paquette RL, Kuriyan J, Sawyers CL. 2002. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell* **2**: 117–125.

Sharma M, Beck AH, Webster JA, Espinosa I, Montgomery K, Varma S, van de Rijn M, Jensen KC, West RB. 2010. Analysis of stromal signatures in the tumor microenvironment of ductal carcinoma in situ. *Breast Cancer Res Treat* **123**: 397–404.

Shipitsin M, Campbell LL, Argani P, Weremowicz S, Bloushtain-Qimron N, Yao J, Nikolskaya T, Serebryiskaya T, Beroukhim R, Hu M, et al. 2007. Molecular definition of breast tumor heterogeneity. *Cancer Cell* **11**: 259–273.

Shirakawa K, Kobayashi H, Heike Y, Kawamoto S, Brechbiel MW, Kasumi F, Iwanaga T, Konishi F, Terada M, Wakasugi H. 2002. Hemodynamics in vasculogenic mimicry and angiogenesis of inflammatory breast cancer xenograft. *Cancer Res* **62**: 560–566.

Shirakawa K, Tsuda H, Heike Y, Kato K, Asada R, Inomata M, Sasaki H, Kasumi F, Yoshimoto M, Iwanaga T, et al. 2001. Absence of endothelial cells, central necrosis, and fibrosis are associated with aggressive inflammatory breast cancer. *Cancer Res* **61**: 445–451.

Shook D, Keller R. 2003. Mechanisms, mechanics and function of epithelial-mesenchymal transitions in early development. *Mech Dev* **120**: 1351–1383.

Siegel R, Ma J, Zou Z, Jemal A. 2014. Cancer statistics, 2014. *CA Cancer J Clin* **64**: 9–29.

Siegel R, Naishadham D, Jemal A. 2013. Cancer statistics, 2013. *CA Cancer J Clin* **63**: 11–30.

Simmons DL, Walker C, Power C, Pigott R. 1990. Molecular cloning of CD31, a putative intercellular adhesion molecule closely related to carcinoembryonic antigen. *J Exp Med* **171**: 2147–2152.

Snuderl M, Fazlollahi L, Le LP, Nitta M, Zhelyazkova BH, Davidson CJ, Akhavanfard S, Cahill DP, Aldape KD, Betensky RA, et al. 2011. Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. *Cancer Cell* **20**: 810–817.

Steeg PS, Theodorescu D. 2008. Metastasis: a therapeutic target for cancer. *Nat Clin Pract Oncol* **5**: 206–219.

Su K-Y, Chen H-Y, Li K-C, Kuo M-L, Yang JC-H, Chan W-K, Ho B-C, Chang G-C, Shih J-Y, Yu S-L, et al. 2012. Pretreatment epidermal growth factor receptor (EGFR) T790M mutation predicts shorter EGFR tyrosine kinase inhibitor response duration in patients with non-small-cell lung cancer. *J Clin Oncol* **30**: 433–440.

Swanton C. 2012. Intratumor heterogeneity: evolution through space and time. *Cancer Res* **72**: 4875–4882.

Szerlip NJ, Pedraza A, Chakravarty D, Azim M, McGuire J, Fang Y, Ozawa T, Holland EC, Huse JT, Jhanwar S, et al. 2012. Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. *Proc Natl Acad Sci USA* **109**: 3041–3046.

Talmadge JE, Fidler IJ. 2010. AACR centennial series: the biology of cancer metastasis: historical perspective. *Cancer Res* **70**: 5649–5669.

Thiery JP. 2002. Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* **2**: 442–454.

Thiery JP, Acloque H, Huang RYJ, Nieto MA. 2009. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**: 871–890.

Tibshirani R, Hastie T, Narasimhan B, Chu G. 2002. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci USA* **99**: 6567–6572.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–515.

Tsai JH, Yang J. 2013. Epithelial-mesenchymal plasticity in carcinoma metastasis. *Genes Dev* **27**: 2192–2206.

Valiente M, Obenauf AC, Jin X, Chen Q, Zhang XH-F, Lee DJ, Chaft JE, Kris MG, Huse JT,

Brogi E, et al. 2014. Serpins promote cancer cell survival and vascular co-option in brain metastasis. *Cell* **156**: 1002–1016.

Vanharanta S, Massagué J. 2013. Origins of metastatic traits. *Cancer Cell* **24**: 410–421.

Vidali C, Caffo O, Aristei C, Bertoni F, Bonetta A, Guenzi M, Iotti C, Leonardi MC, Mussari S, Neri S, et al. 2012. Conservative treatment of breast ductal carcinoma in situ: results of an Italian multi-institutional retrospective study. *Radiat Oncol* **7**: 177.

Virnig BA, Tuttle TM, Shamliyan T, Kane RL. 2010. Ductal carcinoma in situ of the breast: a systematic review of incidence, treatment, and outcomes. *J Natl Cancer Inst* **102**: 170–178.

Weckermann D, Müller P, Wawroschek F, Harzmann R, Riethmüller G, Schlimok G. 2001. Disseminated cytokeratin positive tumor cells in the bone marrow of patients with prostate cancer: detection and prognostic value. *J Urol* **166**: 699–703.

Wiseman BS, Werb Z. 2002. Stromal effects on mammary gland development and breast cancer. *Science* **296**: 1046–1049.

Wu K, Quan Z, Weng Z, Li F, Zhang Y, Yao X, Chen Y, Budman D, Goldberg ID, Shi YE. 2007. Expression of neuronal protein synuclein gamma gene as a novel marker for breast cancer prognosis. *Breast Cancer Res Treat* **101**: 259–267.

Yang J, Mani SA, Donaher JL, Ramaswamy S, Itzykson RA, Come C, Savagner P, Gitelman I, Richardson A, Weinberg RA. 2004. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**: 927–939.

Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, Isakoff SJ, Ciciliano JC, Wells MN, Shah AM, et al. 2013. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339**: 580–584.

Zhang L, Riethdorf S, Wu G, Wang T, Yang K, Peng G, Liu J, Pantel K. 2012. Meta-analysis of the prognostic value of circulating tumor cells in breast cancer. *Clin Cancer Res* **18**: 5701–5710.

Zhou S, Yi T, Zhang B, Huang F, Huang H, Tang J, Zhao X. 2012. Mapping the high throughput SEREX technology screening for novel tumor antigens. *Comb Chem High Throughput Screen* **15**: 202–215.